

Fri Mar 14 14:49:32 2003

us-09-674-237a-3.ra1

Page 1

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:37:52 ; Search time 23 Seconds
(without alignments)
1551.739 Million cell updates/sec

Title: US-09-674-237A-3
Perfect score: 6269
Sequence: 1 MAQFPFGSGSLDVAIVTVE.....VGLFSPNVKLTMDPSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2058.5	32.8	462	4 US-08-630-915A-38	Sequence 38, Appl
2	1388.5	22.1	509	4 US-08-630-915A-194	Sequence 194, App
3	816.5	13.0	248	4 US-08-630-915A-40	Sequence 40, Appl
4	486.5	7.8	896	1 US-08-095-737-2	Sequence 2, Appl
5	486.5	7.8	896	1 US-08-480-145-2	Sequence 2, Appl
6	486.5	7.8	896	2 US-08-477-389-2	Sequence 4, Appl
7	486	7.8	897	1 US-08-095-737-4	Sequence 4, Appl
8	486	7.8	897	1 US-08-480-145-4	Sequence 4, Appl
9	486	7.8	897	2 US-08-477-389-4	Sequence 2, Appl
10	382	6.1	1162	4 US-08-728-333A-2	Sequence 2, Appl
11	382	6.1	1162	4 US-09-298-568-2	Sequence 2, Appl
12	381	6.1	1898	2 US-08-056-200-94	Sequence 94, Appl
13	381	6.1	1898	2 US-08-800-644-94	Sequence 94, Appl
14	333	5.3	553	1 US-08-475-894-2	Sequence 2, Appl
15	333	5.3	553	1 US-08-484-710-2	Sequence 2, Appl
16	333	5.3	553	1 US-08-484-709-2	Sequence 2, Appl
17	333	5.3	553	4 US-08-474-697-2	Sequence 2, Appl
18	325	5.2	659	4 US-08-630-915A-133	Sequence 133, App
19	322	5.1	62	4 US-08-630-915A-135	Sequence 135, App
20	314	5.0	62	4 US-08-630-915A-134	Sequence 134, App
21	309	4.9	56	4 US-08-466-390-4	Sequence 4, Appl
22	297	4.7	2101	1 US-08-470-950-4	Sequence 4, Appl
23	297	4.7	2101	1 US-08-467-781-4	Sequence 4, Appl
24	297	4.7	2101	1 US-08-483-924-4	Sequence 4, Appl
25	297	4.7	2101	2 US-09-452-294-1	Sequence 1, Appl
26	296	4.7	57	4 US-08-630-915A-136	Sequence 136, App
27	296	4.7	57	4 US-08-630-915A-136	Sequence 136, App

28	296	4.7	2101	1 US-08-195-487-4	Sequence 4, Appl
29	296	4.7	2101	5 PCR-US93-06160-4	Sequence 2, Appl
30	281	4.5	1805	1 US-07-853-913-2	Sequence 2, Appl
31	281	4.5	683	6 Patent No. 5210183-3	Sequence 30, Appl
32	281	4.5	788	4 US-08-630-915A-30	Sequence 2, Appl
33	280.5	4.5	1375	4 US-09-722-139-2	Sequence 2, Appl
34	280.5	4.5	1375	4 US-09-721-832-2	Sequence 2, Appl
35	280.5	4.5	1375	4 US-09-393-569-2	Sequence 2, Appl
36	275	4.4	1360	4 US-08-685-871-2	Sequence 2, Appl
37	274.5	4.4	1354	3 US-08-533-306A-2	Sequence 2, Appl
38	274	4.4	576	2 US-08-742-923A-2	Sequence 2, Appl
39	274	4.4	576	2 US-08-533-306A-6	Sequence 6, Appl
40	265.5	4.2	816	2 US-08-742-923A-6	Sequence 6, Appl
41	265.5	4.2	885	2 US-08-533-306A-4	Sequence 4, Appl
42	265.5	4.2	885	2 US-08-742-923A-4	Sequence 4, Appl
43	263	4.1	62	4 US-08-630-915A-138	Sequence 138, App
44	263	4.1	955	1 US-08-006-676B-1	Sequence 1, Appl
45	259.5	4.1	955	1 US-08-006-676B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-630-915A-38
Sequence 38, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLER, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-38
Query Match 32.8%; Score 2058.5; DB 4; Length 462;
Best Local Similarity 79.3%; Pred. No. 2e-138;
Matches 391; Conservative 13; Mismatches 18; Indels 71; Gaps 1;

QY 721 PMSTTEKPLTISAQESKVVYVYALYFESRSHDEITTOPGDIYVWDESGTGEWLG 780
 Db 41 PGPLOKGPPLTISAQENKVVYVYALYFESRSHDEITTOPGDIYVWDESGTGEWLG 100
 QY 781 ELKKTGFPANVAKIPENEVPAPKPVYDLTSAPAKLARETPAPLPTVSSTTP 840
 Db 101 ELKKTGFPANVAKIPENEVPAPKPVYDLTSAPAKLARETPAPLPTVSSTTP 160
 QY 841 NNMADPSSSTSSNEKPELTDNDWTAAPSLTVPASQFORSAFTPATATSSSPVL 900
 Db 161 NNMADPSSSTSSNEKPELTDNDWTAAPSLTVPASQFORSAFTPATATSSSPVL 220
 QY 901 GQGEKEVGLQAQALYPMRAKNDHNFNSKSDYITVLEQDDMMWGEVQOGKMPKSYK 960
 Db 221 GQGEKEVGLQAQALYPMRAKNDHNFNSKSDYITVLEQDDMMWGEVQOGKMPKSYK 280
 QY 961 LISGPVKSTIDTGTSPASLKRVAAPAKPALPGEEFIAMTYESSEGDLTFQGD 1020
 Db 281 LISGPVKSTIDTGTSPASLKRVAAPAKPALPGEEFIAMTYESSEGDLTFQGD 318
 QY 1021 VIVVTKKDGMWGTGVGSGVFPNRYRLMDSESGTAGTSLGKKPIAVIASYAA 1080
 Db 319 -----EIAQVIASYAA 329
 QY 1081 TGPEOLTLAPGOLIRKKNPGWMEGLQARKKRQIGFNPANYVLLSPGSKITPTE 1140
 Db 330 TGPEOLTLAPGOLIRKKNPGWMEGLQARKKRQIGFNPANYVLLSPGSKITPTE 389
 QY 1141 LPTAVOPAVCOVIGMYDYTAQNDDELAFSGQIINVLNKEDPMWKGVSQVGLFPSN 1200
 Db 390 LPTAVOPAVCOVIGMYDYTAQNDDELAFSGQIINVLNKEDPMWKGVSQVGLFPSN 449
 QY 1201 YVKLTMDPSQO 1213
 Db 450 YVKLTMDPSQO 462

RESULT 2

US-08-630-915A-194

Sequence 194, Application US/08630915A
 Patent No. 6309820

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
 APPLICANT: HOFFMAN, No. 6309820h
 APPLICANT: KAY, Brian K.
 APPLICANT: FOWLES, Dana M.
 APPLICANT: MCCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/630,915A
 FILING DATE: 03-APR-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 194:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 509 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-630-915A-194

Query Match

Best Local Similarity 22.1%; Score 138.5; DB 4; Length 509;

Matches 287; Conservative 85; Mismatches 124; Indels 91; Gaps 14;

QY 635 ERKSELEKOK-----EPAQRVQERDKOMLEHVQOEOPRPRKHEEDRLKREDSYKKE 690
 Db 6 ERKSELEKOKLEBPAKAKAKOGKKNLWKNLKEE-----EKOKRIQEKTEQKIQ 59
 QY 691 AERRAPENQDKOSRLEPHHOEPAPKATQAPWSTTEKPLTISAQESVYVYVYALYYPE 750
 Db 60 EERRAKKEKQ-----RETASV-----LVNRYALYPE 86
 QY 751 SRSHDEITTOPGDIYVWDESGTGEWLGELKKTGFPANVAKIPE-ENEVPAPK 808
 Db 87 ARNHDEMSFNSGDIQVDEKTVGEPGLYSGFQGNFGPCNVYKMPSENE----- 139
 QY 809 VVDLSAPAPKALARETPAPLPTVSSTTPNNMADFSTWSSSEKPELTDNDWTA 868
 Db 140 -----KAVSPKAL-----LPTVLSLATS-----TSSEPLSSNOASVTDYQ- VS 180
 QY 869 QPSLTVPASQFORSAFTPATATGSSSPVLGQGEKEVGLQAQALYPMRAKNDHNFN 928
 Db 181 FSNLTIVTSMQ-KKSAFTRTVSPG-SVSPHQQGVENLKAQALCSWAKAKDHLNFS 237
 QY 929 KSDYITVLEQDDMMWGEVQOGKMPKSYVYKLSGPVKSTSIDTGTSPASLKRVA 988
 Db 238 KHDITVLEQDDMMWGEVQOGKMPKSYVYKILPGSEK-----REEPALVAAVN 289
 QY 989 --PAKPAIPEEFIAMTYESSEGDLTFQGDVIVYTKKDGDMWGTGVGSGVFPN 1046
 Db 290 KKPTSAVYSGEYIALYVSSVERDLPFTGEBILVTQKDGDMWGTGVGSGVFPN 349
 QY 1047 YVRLKDSGSGTGTGKTSIGKKPEIAQVYASYATGPEOLTLAPGOLIRKKNPGWME 1106
 Db 350 YVRLKDSGSGTGTGKTSIGKKPEIAQVYASYATGPEOLTLAPGOLIRKKNPGWME 409
 QY 1107 GELQARKKRQIGFNPANYVLLSPGSKITPTELPKTAQVOPAVCOVIGMYDYTAQND 1166
 Db 410 GELQARKKRQIGFNPANYVLLSPGSKITPTELPKTAQVOPAVCOVIGMYDYTAQND 462
 QY 1167 LAFSGQIINVLNKEDPMWKGVSQVGLFPSNYVLLTMDPSQO 1213
 Db 463 LAFSGQIINVLNKEDPMWKGVSQVGLFPSNYVLLTMDPSQO 509

RESULT 3

US-08-630-915A-40

Sequence 40, Application US/08630915A
 Patent No. 6309820

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
 APPLICANT: HOFFMAN, No. 6309820h
 APPLICANT: KAY, Brian K.
 APPLICANT: FOWLES, Dana M.
 APPLICANT: MCCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP

```

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-40

```

```

Query Match 13.0%; Score 816.5; DB 4; Length 248;
Best Local Similarity 58.4%; Pred. No. 1.6e-50; Indels 17; Gaps 3;
Matches 153; Conservative 38; Mismatches 54;

```

```

QY 954 FPKSVYKLSIGFVRKSTIDTPTSPASLKRVAS--PAKPAIPEGEEFIAMYTESESQ 1011
DB 2 FFAATKTIIPGEVYK-----REPEALVAANVKRPTSAAYSVGEYIALYPSSEVP 53
QY 1012 GDLTFQGDVIVTKKDGDMWTGTVGDKSGVFPSSNYVRLKSESGTGKGLSKKEI 1071
DB 54 GDLTFEGEILIVTKDDEWMTGSIIGDRSGIFPSNYVVRKPOESFGASKSGASNKKEI 113
QY 1072 AQVIAVYATGEQELTLARGLILIRKKNPGWEGELQARGKKQIQMFNPANYKLSP 1131
DB 114 AQVIAVYASGEQSLAPGQILILKKNVSGWQELQARGKKQKQMFPAHVKLGP 173
QY 1132 GSKITPTLPTAVQAVQVIGWYDYTAQNDDELAFSGQIIVLNKEDPDWVGES 1191
DB 174 SSKRATPAHP-----VCQVIANYDYAANNDELSFSKQGLIVNKKDDPDWVGESIN 226
QY 1192 GQVGLFPPSNVYKLTDDMPDQQ 1213
DB 227 GVTGLFPPSNVYKMTDSDPSQQ 248

```

```

RESULT 4
US-08-095-737-2
Sequence 2, Application US/08095737
Patent No. 5487979
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,737
FILING DATE: 19930722
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 896 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-737-2

```

```

Query Match 7.8%; Score 486.5; DB 1; Length 896;
Best Local Similarity 20.5%; Pred. No. 2.9e-26;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

```

```

QY 15 WATTEERAKHDOQFLSLKPIGFTTQDARNFPPQSLPQVLAQIYALADNNDRMD 74
DB 122 WAKPEDKAKYDAIFDLSLPVNGFSGDKVYKVLNSKLPVILGRWELSDIDHMD 181
QY 75 QVESIAMKLIKLLQGYOLPSTLPVWKQOPVAISSAPAFIGIGIASNPITAAVAPVM 134
DB 182 RDEFVAMFLVYALE-----XEPVPM 203
QY 135 GSIPVGMGEPVSVPPAAPPVPLANGAPPVLOPLPAFAHPAATPMPKSSFSRSGSQL 194
DB 204 -----SLPPALVPP-----SKR 215
QY 195 NTYKQAQSFVVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHLTGPQARTIMQSSL 254
DB 216 KT-----WVSPAEKAKYDEIFLTKDKMDGFPVSGLEVAFILFKTGL 257
QY 255 POAQLASIMNLSIDIDQKLTAEFTLAMHLIDVAN--SQQLPVPVLPPEYIPPSERRVS 313
DB 258 PSTLAIHWLSCOTKCKGKLSKQDFALAPHLISQKLIGIDPHVLTPEMIPPS---DPA 314
QY 314 GSGNSVSISSSVQORLPPEPSSDEQQPEKLPVTFEDKKREFFERGVSLERRQALLE 373
DB 315 SLOKNIITGSSPV-----ADFSAIK-----ELDTLNNETVD 344
QY 374 QQRKEQRLAQLERAEQERERQEQEAKRQLELEKQLEKQRELEPROBEERKEIER 433
DB 345 LQREK-----NNVEDDLKEKE-----DTIKQTSVQ---DLQDEVQRE 380
QY 434 EAAKRELEQROLEMERNRROELLNKNKEQEGTVVLKARKKTLELEALNDKXQLEB 493
DB 381 NTNQKIQAKQKQ-----QVQELLDE-----LDQKQALEB 410
QY 494 KLQDIRCLATQROEISTKSKBELRLAETHLOQLOQESQOMGLRILPEKQLISQQLK 553
DB 411 QLEKVRKCAEQAQLISSLK-----AELTQESQISTYEEELAK-----AREELSR 456
QY 554 VQONSLSRDSLTLKRALKEKELARQOLRQOLDEVERETRS-----KLOEIDVFNQOLK 607
DB 457 LQQETAE-----LEESVSGKQALEPLQGHLDQSQEISMQMKLMEMKDLNHNHSQLN 510
QY 608 ELREIHS-----KQQLQKQSLFANRLKQEQERKLEL-----641
DB 511 WCSPHSLIVNGATDYCSLSTSSSETANLNHVEGQNLSEPIHOSPPARSSPELLPSG 570
QY 642 EKQKEDAKRRVQERDKQWLEHVQOEQEPRPKRPHEDRLKREDSVRKKEAER-----694

```

Db 571 VTDENEVTATVEKVCSELDN-----NHSKEEDPFVNDSSSLTGPAVDNLDFQGS 622
 Qy 695 -----AKPEMODKOSRLFHPHOEPK-----LATOAPWSTTEKPLTISAQESV 738
 Db 623 DPFVGSDFPKDDPFGKIDPFGGDPFGKSDPFASDCFFRQSTDPFATSSTDPFSAANSSI 682
 Qy 739 KVVYRRLYPFESRSHDEITIQPGDIVMDESOT-----GEGWLGC-----780
 Db 683 TSV-----ETLKHNDPFAPGGTVVAASDASATDPFASVFGNESFCGGFADPSTLSKVN 734
 Qy 781 ---ELKGTGMPFAN-----YAEKIPENEVPTPAKPVVDLTSAPKALARETAPLP 830
 Db 735 NEDPFRSATSSSVSNVITKNVFEETSVKSEDEPPALP-----PKIGTPTRCPPLP 785
 Qy 831 -----VTSSEP-----STPNMADFSSTWSSSNEKETDNMDT 865
 Db 786 PGKRSINKLSDPDPFKLNDPFGPPGNDSPKEXDPEMFCDPFTSATTTTNKADPSPNFAN 845
 Qy 866 WAAQPS 871
 Db 846 FSAVPS 851

RESULT 5
 US-08-480-145-2
 Sequence 2, Application US/08480145
 Patent No. 5717067
 GENERAL INFORMATION:
 APPLICANT: DiFiore, Pier P
 APPLICANT: Fazioli, Francesca
 TITLE OF INVENTION: A Substrate for the Epidermal Growth
 TITLE OF INVENTION: Factor Receptor Kinase
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center Drive, Sixteenth Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: United States of America
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,145
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/095,737
 FILING DATE: 22-JUL-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelien, Ned A
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH060.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 896 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-480-145-2

Query Match 7.8%; Score 486.5; DB 1; Length 896;
 Best Local Similarity 20.5%; Pred. No. 2,9e-26;
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;
 Qy 15 WAITVEERAKHQOQLSLKPIAGFITGDQARNPFQSGILPQVLAQIWAALDMNNDGMD 74

Db 122 WAVEKPEDAKKDALPDSLSPVNGFLSGDKVXVPVLNLSKLPVDILGRWELSDIDHDMLD 181
 Qy 75 QVEFSIANKLILKILKQIGQLPSTLPVWKQOPVAISSAPAFGIGIASMPPLTAVAPVM 134
 Db 182 RDEFVAMFVLVYCALE-----KEEVPVM 203
 Qy 135 GSIPVWGSPLVSSVPAAVPLANGAPVIGPLPAFAHPAATWPKSSSFRSGSQSL 194
 Db 204 -----SLPPALVPP-----SKR 215
 Qy 195 NTKQAKQSPVVASAPPAEWAAPQSSRLKXRLQFNSHDKTMSGHLTGPOARTILMQSSL 254
 Db 216 XT-----WVVSAPAKAKYDEIFLTKDMQDVGSLEREIFLTKGL 257
 Qy 255 PQOALASITWNLSDIDQCKLTAEEFILAMHLIDVAM--SGQPLPVLPPEXIPPSFRVRS 313
 Db 258 PSTLAAHIMSLCTDKCKGSKSDQFALAFHLISQKLLKGIIDPPHVLTPEMI PPS--DRA 314
 Qy 314 GSGMSVSSSSVDQRLDEPSSSEDOQPEKKLPYTFEDKKRENERGSEVELERRQALLE 373
 Db 315 SLQKNIIIGSPV-----ADFGAIK-----ELDTLNNEIVD 344
 Qy 374 QQRKEQERLADLERAEQERKERERQEOBAKQOLEBLEKOLEKORELERQREERKETEIR 433
 Db 345 LQREK-----NNVEDDLKE-----DTIKORSEVO--DLODEVORE 380
 Qy 434 EAAKRELERQRLERWERNRROELNQRNKEOGTVVKARKRTLEFLEALANKKQLEG 493
 Db 381 NTNIGKQAOQO-----QOVELLDE-----LDQKQALE 410
 Qy 494 KLQDIRCLATQROELISTNKSRELRJAETTHLQOQLOESQOMGLRIIPKQILSDQLK 553
 Db 411 QLKEVRKCAEAQILISLK-----AELTQSOISQISTYEELIAK-----AREELSR 456
 Qy 554 VQOQSLRDSLLTLKRLAEKELARQOREQDEVERETS-----KLQELIVFNNOJK 607
 Db 457 LQOETAE-----LEESVESGKAQLEPLQOHLDSQOELISSMQKLMEMKDLNHNSQLN 510
 Qy 608 ELREIHS-----KQLOKQSLAEARLKQEKORKSLEL-----641
 Db 511 WCSSPHSILVNGATDYCSLSTSSSETANLNEHVGOSNLESEPIHQESPARSPPELLPSG 570
 Qy 642 EKQEDAKQRYQERDKQWLHVQOEEQPPRKRKHEDRLKREDSVKKEAEER-----694
 Db 571 VTDENEVTATVEKVCSELDN-----NHSKEEDPFVNDSSSLTGPAVDNLDFQGS 622
 Qy 695 -----AKPEMODKOSRLFHPHOEPK-----LATOAPWSTTEKPLTISAQESV 738
 Db 623 DPFVGSDFPKDDPFGKIDPFGGDPFGKSDPFASDCFFRQSTDPFATSSTDPFSAANSSI 682
 Qy 739 KVVYRRLYPFESRSHDEITIQPGDIVMDESOT-----GEGWLGC-----780
 Db 683 TSV-----ETLKHNDPFAPGGTVVAASDASATDPFASVFGNESFCGGFADPSTLSKVN 734
 Qy 781 ---ELKGTGMPFAN-----YAEKIPENEVPTPAKPVVDLTSAPKALARETAPLP 830
 Db 735 NEDPFRSATSSSVSNVITKNVFEETSVKSEDEPPALP-----PKIGTPTRCPPLP 785
 Qy 831 -----VTSSEP-----STPNMADFSSTWSSSNEKETDNMDT 865
 Db 786 PGKRSINKLSDPDPFKLNDPFGPPGNDSPKEXDPEMFCDPFTSATTTTNKADPSPNFAN 845
 Qy 866 WAAQPS 871
 Db 846 FSAVPS 851

RESULT 6
 US-08-477-389-2
 Sequence 2, Application US/08477389
 Patent No. 5872219
 GENERAL INFORMATION:

APPLICANT: DiFiore, Pier P
 APPLICANT: Fazioli, Francesca
 TITLE OF INVENTION: A Substrate for the Epidermal Growth
 FACTOR RECEPTOR KINASE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center Drive, Sixteenth Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: United States of America
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,389
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/095,737
 FILING DATE: 22-JUL-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelsen, Ned A.
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH060.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 896 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-477-389-2

Query Match 7.88; Score 486.5; DB 2; Length 896;
 Best local similarity 20.58; Pred. No. 2,9e-26;
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;
 QY 15 WAIVBRKADQOFLSKPIAGITGQDARNFFQSGLPQPVLAQIMALAMNNDGRMD 74
 DB 122 WAWPREDAKDAIFDSIPVNGFLSGDKKVPVLLNKTLPVILIGRWELSDIDHDMLD 181
 QY 75 QVESIAMKLIKLIKQGYLPSTLPPVKKQPVAISSAPAFGIGIASMPPLTVAVPVM 134
 DB 182 RDEFVAMFLVYCALE-----KEFVPM 203
 QY 135 GSIPVMSPLVSVPPAANVPPLANGAPVPIQPLPAFAHPAATWPKSSSFSGPSQL 194
 DB 204 -----SLPPLALVP-----SKR 215
 QY 195 NTKLOKQSFVVASPPAAEWAPOSRLKYROLFNSHDKTMSGHLTGQARTILMOSL 254
 DB 216 KT-----WVSPAELAKYDEIFLKTDXMDGVSLLEVRFLFKTGL 257
 QY 255 PQOALASIMNLSIDDDQKLTAEFTILAMHLIDVAM-SGQPLPVLPPEYIPPSFRFRRS 313
 DB 258 PSTLIAHIMSLCDTKDCGKLSKQDFALAFHLISQKLIKIDPPHVLTPEMIPPS---DRA 314
 QY 314 GSGMVISSSSVQDLPPPESSDEQOEKKLPVTEEDKKRNFERSGVELEKROALLE 373
 DB 315 SLQKNITGSSPV-----ADFSATK-----ELDTLANEIVD 344
 QY 374 QORKEORLAQLERABOERKEREROEAKROLEKOLEKORELEROREERKEIERR 433
 DB 345 LQREK-----NNVEQDLKEKE-----DTIKQRTSEVO---DLQDEVQRE 380
 QY 434 EAKKELEFROLEMERNRROELLAMORNKKEGCTVVLAKRRTLEFLEALNDKHOLEG 493

DB 381 NTNLQKLOAQOKO-----QVQELDE-----LDFOKAQLE 410
 QY 494 KLODIRCLATOROEREISTNKSRELRIATITLCOOLOESQOMGRILPEKQILSDQKQ 553
 DB 411 QKEVRKKCAEBAQILSLK-----AELTSQESQISTEELAK-----ARELSR 456
 QY 554 VQONSLHSDSLTLKRALAEKARQOLREQDVEVERFTR-----KLOEIVFNQK 607
 DB 457 LQETPAE-----LEESVESGKAQLEPQOHLQDPSQOQISMMQKMLMEMDLNHNQNL 510
 QY 608 ELRETHS-----KQLOKORSLEAARLKQKQERKSUEL----- 641
 DB 511 WCSSPHSILVNGATDYCSLSTSSSETANLNEHVGSNLSBPTHQSSPARSSPELLPSG 570
 QY 642 EKQKEDAQROVERDKQMLEHVQOEOPRPKPHEDRLKREDSVRKKEAER----- 694
 DB 571 VTDEHEVTAATEKVCSSLN-----NNHKSKEEDPFWNDSSLGVPADTNLDFFQS 622
 QY 695 -----AKPEWODKOSRLFPHQEPK-----LATQAPMSTTEKGLTISAOESV 738
 DB 623 DPFVGSDFPKDDPFQKIDPFQGDPEKGSDFPASDCFRQSTDPFATSTDPFSAANSSI 682
 QY 739 KVVYRRLVPEESRSHDITTOPGDIVWVDESOT-----GEGWLGQ----- 780
 DB 683 TSV-----ETLKHNDPPAPGGIVVAASDATDPFASVFGNSFGGFGFADFTLSKVN 734
 QY 781 ---ELKGTGWFPAN-----YAEKIPENEVPTPAKEVTDLTSAAPAKLARETPALP 830
 DB 735 NDDPRSATSSSVSNVVTQKVFEETSVKSEDEPALP-----PKIGTPTRPCPL 785
 QY 831 -----VTSEEP-----STPNMADFSSTWPSNNKPKPTDMDT 865
 DB 786 PGKRSINKLSDSPFPKLNDDPQPPGNDSPKQPEMFCDFTSATTITNKADPSNPNAN 845
 QY 866 MAAPS 871
 DB 846 FSAYPS 851

RESULT 7
 US-08-095-737-4
 Sequence 4, Application US/08095737
 Patent No. 5487979
 GENERAL INFORMATION:
 APPLICANT: DiFiore, Pier P
 APPLICANT: Fazioli, Francesca
 TITLE OF INVENTION: A Substrate for the Epidermal Growth
 FACTOR RECEPTOR KINASE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center Drive, Sixteenth Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: United States of America
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/095,737
 FILING DATE: 19930722
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelsen, Ned A.
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH060.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
 LENGTH: 897 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-095-737-4

Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 3.2e-26;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

QY 15 WATTEERAKHDQOFLSLKPIAGFITGDQARNFFQSGLPQVLAQIMALAMNNDGRND 74
DB 122 WAKSSEKAKYDAIFPSLSVDGFLSGDKVKVPLNSKLPVELIGKRWELSDIDHDKLD 181
QY 75 QVEFSIAMKLIKLIKLOGYQLPSTLPVPMKQPVVAISSAPAFGIGIASMPPLTAAVAVPM 134
DB 182 RDEFVAMFLVYCALE-----KEPVPM 203
QY 135 GSIPVVGMSBPPLVSVPPAAVPPPLANGAPPVIOPLPAFAHPAATWPKSSFSRSGPSQL 194
DB 204 -----SLPPLVPP-----SKR 215
QY 195 NTKLQAKQSFVVASAPAEWAVPOSSRLKYROLFNSHDKTMSGHLTGPARTILMOSGL 254
DB 216 KT-----WVSPAEKAKYDEIFLKTDKMDGYVSGLEVETFLKTGL 257
QY 255 PQOQLASINWLSIDDDGKLTAEFLMHLIDVAM-SGQPLPVLPPRYIPPSFRVRS 313
DB 258 PSALLHIMSLCTCGKGLSKQFALAPHLINOKLIKIDPHSITPMPIPPS--DRS 314
QY 314 GSGMSYISSVDQRLPEPPSSDEQPEPKLPVTFEDKKRENFRGSLTEKRRQALL 373
DB 315 SLQKNTGSSPV-----ADPSAIK-----ELDTLNNEIVD 344
QY 374 QQRKEQERLAQLERAOEKEREEROEAKROLLEKOLEKORELEKOREEERKEIERR 433
DB 345 LOREK-----NNVEQDLKEKE-----DTVKQRTSEVQ--DLQOEVOQRE 380
QY 434 EAKARELEROROLEMERNRNROELNORNKEGEGTVLTKARKRTLEFELEALNDKHOLEG 493
DB 381 SINLQKLOKQKQ-----OVQELGEG-----LDEQKQALEE 410
QY 494 KLQDIRCLATOROELSTNKSRELRIAEITHLQOOLQESQOMLGRLLPEKQILSDQKQ 553
DB 411 QLOEYAKKCAEEAQLISLQ-----AEITQESQISSYEELK-----AAEELSR 456
QY 554 VQONSILHRSLTLTKALEAKELARQQLREQULDEVRE-----TRSKIQEIDVFNQULKE 608
DB 457 LQOETTAQ-----LEESVESGKAQLEPLQOHLQESQOEISSQMRLKEMKDETENNQ-- 507
QY 609 LREHSKQOLQKQ-----RSLBAARLKQKEQERKSLLEKQKEDANQORVQ 653
DB 508 -----SNMSSSPQSVLVNGATDYCSLSTSSSTANFNENAEONNLESEPHIQESVSS 562
QY 654 ER-----DKOMLEHVQOEQPRPKKHEEDRLKREDSVRKKEA-----EERAKPEMDK 702
DB 563 PEIAPSDVYDESEAVTVAGNEKVTPR--FDDDKHSEKEDPEFVSSSLTDAVDATNLDLF 620
QY 703 GSRLLF-----HPHOEPALTAQAPWSTTEKPELITSQESKVVVY 743
DB 621 QSDPFGVSGPFDKDPFGKIDPFPGSDPFKSDPFASDCFFKQISTDPFTTSTDPSPASNN 680
QY 744 RALYPFEESHDEITIPQDVIWVDESQTEGP--GMLGELKGTGWPAPVAYAEKIPENE 801
DB 681 SNTSVEETWKHNDPPAPGGTVVAAASDSATDPFASVFNQESFG--DGDAFPTLSKVNND 739
QY 802 VETP-----AKPYTDLTSA-----PA--PKALAEETPAPLPYTSSPSSTPPNN 842
DB 740 AENPTISSTSVTLAKPVELETASKSEDVPPALPPKGTPTRPCPP-----PGKREIN 794
QY 843 WADFSST-----WPSSSNEKRETDNMDTMAAQPSLTVPSAGQLRQSAFTPTATQSS 895

```

DB 795 KLDSSDLKINDPFPQPPGNDSPKXKPDWFC-----DPFTSTTNKE 838
 QY 896 PSP 898
 DB 839 ADP 841

RESULT 8

```

US-08-480-145-4
: Sequence 4, Application US/08480145
: Patent No. 571067
: GENERAL INFORMATION:
: APPLICANT: Diflore, Pier P
: APPLICANT: Fazio, Francesca
: TITLE OF INVENTION: A Substrate for the Epidermal Growth
: TITLE OF INVENTION: Factor Receptor Kinase
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: United States of America
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,145
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 530
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/095,737
: FILING DATE: 22-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned A
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH060.001A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 897 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-480-145-4

```

Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 3.2e-26;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

QY 15 WATTEERAKHDQOFLSLKPIAGFITGDQARNFFQSGLPQVLAQIMALAMNNDGRND 74
DB 122 WAKSSEKAKYDAIFPSLSVDGFLSGDKVKVPLNSKLPVELIGKRWELSDIDHDKLD 181
QY 75 QVEFSIAMKLIKLIKLOGYQLPSTLPVPMKQPVVAISSAPAFGIGIASMPPLTAAVAVPM 134
DB 182 RDEFVAMFLVYCALE-----KEPVPM 203
QY 135 GSIPVVGMSBPPLVSVPPAAVPPPLANGAPPVIOPLPAFAHPAATWPKSSFSRSGPSQL 194
DB 204 -----SLPPLVPP-----SKR 215
QY 195 NTKLQAKQSFVVASAPAEWAVPOSSRLKYROLFNSHDKTMSGHLTGPQARTILMOSGL 254
DB 216 KT-----WVSPAEKAKYDEIFLKTDKMDGYVSGLEVETFLKTGL 257
QY 255 PQOQLASINWLSIDDDGKLTAEFLMHLIDVAM-SGQPLPVLPPRYIPPSFRVRS 313

```

```

Db 258 PSALLAHISLCTKCGKLSKQDFALAFHLINOKLIKIDDPHSITPMPPS---DRS 314
Qy 314 GSGMSVSSSVYDRLPREPSSSEDEQPEKLVPTFEDKRENFERSVELEKRRQALLE 373
Db 315 SLQKNITGSSPV-----ADFSAIK-----ELDTLNNEIYD 344
Qy 374 QQRKEORLAQLERAEQERKERERQOEAKRQLELEKOLEKORELERQREERKEIER 433
Db 345 LQREK-----NNVEQDLKEKE-----DTVKQRTSEVQ---DLODEVORE 380
Qy 434 EAAKRELERQRELEWERNRROELNQRNKEQEGTVLVKARKTLEFELALNDKKHOLEG 493
Db 381 SINQLQLOAQKQ-----QVQELLGE-----LDEQKQALEE 410
Qy 494 KLQDICRLATQROEISTKSKRELRIAEITHLQOOLQESQOMLGRILPEKQILSDQLOK 553
Db 411 QLOEVKCKCAEEAQLISSLK-----AETTSQESQISSYEELK-----AREELSR 456
Qy 554 VQONSILHRDSLTLTKRALEAKELARQOLREOLDEVERE-----TRSKLOEIDVFNQOLKE 608
Db 457 LQOETAQ-----LEESVESGKAQLEPLQOHLQESQOELISSMQRLKMLKDLTDNNQ--- 507
Qy 609 LREIHSKQOLOKQ-----RSLAARLKQKEQERKSLLELEKQEDAQRRVQ 653
Db 508 -----SNWSSSPQSVLVNGATDYCSLSTSSSETANFNENAGQNNLESEPTHOESSVRSS 562
Qy 654 ER-----DKQMLEHYQOEOPRPKRKHEEDRLKREDSVRKKEA-----EERAKPEMDQK 702
Db 563 PELAPSDVTDESAVTVAGNEKVTPR--FDDDKHSKEEDPFNVSSSLDVAADTNLDF 620
Qy 703 QSRLEF-----HPHOEPAKLATQAPWSTTEKGPULTISAQESVKVYVY 743
Db 621 QSDPFGVQSDPFKDDPFKIDPFQGDPPKSDPPASDCFFKQSTSTDPFTSSITDPFSASSN 680
Qy 744 RALYPRESRHDELTTPQDGIWVWVDESQTEP--GMLGSLKQKTMFPANVYKIPENH 801
Db 681 SSNTSVETWGHNDPFAFGTVVAAASDASATDPPASVFGNSFG--DGFADFSTLSKVNND 739
Qy 802 VPRP-----AKPVTDLTSA-----PA--PKLALRETPAPLPVTSSEPTTPNN 842
Db 740 ANFPPTISSSTSSVTIAKPMLEETASKSEDPALPKPVGTPTPCPP-----PGKRPIN 794
Qy 843 WADFSST-----WPSSSNEKETDNDMTMAAOPSLTVPSAQOLRQRAFTPATAGSS 895
Db 795 KLSQSDPLKINDPFPQRPGRNDSPKXDPDMFC-----DPTSTSTTNKE 838
Qy 896 PSP 898
Db 839 ADP 841

```

```

RESULT 9
US-08-477-389-4
; Sequence 4, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,389
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-389-4

```

```

Query Match 7.8%; Score 486; DB 2; Length 897;
Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

```

Qy 15 WAITVEERAKHDOOFLSKPIAGITGDQARNFPFQSGLPQPVLAQI WALADNNDRMD 74
Db 122 WAVSEDAKAKDAIFDSLSPVDSGLSGKVPVVLNKGKLPVEILIGRWELSDIDHDKLD 181
Qy 75 QVESIAMKLIKILKQYQLPSTLPVWKQOPVAISSAPAFGIGIASMPPLTVAAYPM 134
Db 182 RDEFVAMFLVYCALE-----KEEVPVM 203
Qy 135 GSIPVVGSPPLVSSVPPAAYPLANGAPVITQPLPAFHPAATWPKSSSPSGSQOL 194
Db 204 -----SLPPLALVP-----SKR 215
Qy 195 NTKLOKASFDVASAPPAEWAVPQSSRLKYROLFNHDKTMSGHLTGPQARTILMOSL 254
Db 216 KT-----WVSGPAEKAKYDELFLKTDKMDGYVGLRETFELTKGL 257
Qy 255 PQQLASITWNLSDIDQDKLTAEFFILAMHLIDVAM--SGQPLPVLPPEVTPSPFRYRS 313
Db 258 PSALLAHISLCTKCGKLSKQDFALAFHLINOKLIKIDPPHSITPMPPS---DRS 314
Qy 314 GSGMSVSSSVYDRLPREPSSSEDEQPEKLVPTFEDKRENFERSVELEKRRQALLE 373
Db 315 SLQKNITGSSPV-----ADFSAIK-----ELDTLNNEIYD 344
Qy 374 QQRKEORLAQLERAEQERKERERQOEAKRQLELEKOLEKORELERQREERKEIER 433
Db 345 LQREK-----NNVEQDLKEKE-----DTVKQRTSEVQ---DLODEVORE 380
Qy 434 EAAKRELERQRELEWERNRROELNQRNKEQEGTVLVKARKTLEFELALNDKKHOLEG 493
Db 381 SINQLQLOAQKQ-----QVQELLGE-----LDEQKQALEE 410
Qy 494 KLQDICRLATQROEISTKSKRELRIAEITHLQOOLQESQOMLGRILPEKQILSDQLOK 553
Db 411 QLOEVKCKCAEEAQLISSLK-----AETTSQESQISSYEELK-----AREELSR 456
Qy 554 VQONSILHRDSLTLTKRALEAKELARQOLREOLDEVERE-----TRSKLOEIDVFNQOLKE 608
Db 457 LQOETAQ-----LEESVESGKAQLEPLQOHLQESQOELISSMQRLKMLKDLTDNNQ--- 507
Qy 609 LREIHSKQOLOKQ-----RSLAARLKQKEQERKSLLELEKQEDAQRRVQ 653
Db 508 -----SNWSSSPQSVLVNGATDYCSLSTSSSETANFNENAGQNNLESEPTHOESSVRSS 562
Qy 654 ER-----DKQMLEHYQOEOPRPKRKHEEDRLKREDSVRKKEA-----EERAKPEMDK 702
Db 563 PELAPSDVTDESAVTVAGNEKVTPR--FDDDKHSKEEDPFNVSSSLDVAADTNLDF 620

```


REFERENCE/DOCKET NUMBER: NIH054.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (714) 760-0404
 TELEFAX: (714) 760-9502
 INFORMATION FOR SEQ ID NO: 94:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1898 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-800-644-94

Query Match 6.1%; Score 381; DB 2; Length 1898;
 Best Local Similarity 27.0%; Pred. No. 2.7e-18;
 Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

326 DRLPPESEDEEQPEKLPVTFEDKKRNPFERSVLEKRRQALLLEQ--RKEOE--R 381
 348 EQQLRR--QEERRRQQLRREGEEERRRQQLRREGEEERRRQQLRREGQQLR 404
 382 LAQLERAQERKERE--RQEDAKR--OLELEKQLEKQLEKQLEKQLEKQLEKQ 426
 405 EQQLRRQQLRREGQQLRREGQQLRREGQQLRREGQQLRREGQQLRREGQQLR 464
 427 ---RKEIEREAKRLEERORQLEWENRRQELNQNKEQEGTVLAKARKTLELEA 483
 465 RLKREQERRDWLKREETERH--EOERRQQL--KQDQEE--RRRWLKLLEE- 513
 484 LNDKQLEKGLDQIRCLATQROEISTNKSRLIAETTHLQOOLQESQ--QML- 537
 514 -ERREQERREQQLR--REQERREQQLRQEEERLQQLRSEQQLRRQERRLQLLK 570
 538 -----GRLLPEKQLSDQ---KQVQNSLRHSDSLTLTKRLAKELARQ 580
 571 REBEKRLQERRRQQLRQEEERDQQLKEEERQQLRQEEERLQKEEERLE 630
 581 LREOLDE-VERETSKLQELIDVENNQLKELR---EIKSKQQLKQSL---EARKLQK 632
 631 QEBERDRRLKREBEERREHHLKSEQERRHQQLRREGQERRLKREBEERLEQ 690
 633 ---EOERSLLEL-EKQEDAQRRVOERDKQLEHVOEEQPPRK----- 673
 691 LKREHEERREOELEAEEOQARERIKSRIPKQWQLESADARQSKVLLEAPQARABA 750
 674 PHEEDRLKRDYVKKKAEERAKPEMODKQSLFPHQBEAKLATAPMSTTEKGLTS 733
 751 PQDEERKRESELQWQEEERARHQDEEQRKDFTWQOAE-----EKSEGRQRLS 803
 734 ADESXVYVYRATVPFESRSHDE 756
 804 APPLRQERROLRAERQREQ 826

RESULT 14
 US-08-475-894-2
 Sequence 2, Application US/08475894
 Patent No. 5641748
 GENERAL INFORMATION:
 APPLICANT: Yen-Ming Hsu
 TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 State Street, Suite 510
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/475,894
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Louis Myers
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: BGP-191
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 553 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 US-08-475-894-2

Query Match 5.3%; Score 333; DB 1; Length 553;
 Best Local Similarity 27.4%; Pred. No. 1.2e-15;
 Matches 93; Conservative 67; Mismatches 115; Indels 64; Gaps 15;

912 QALYPMARKDNHLENKSDVITVLEQD--MWFGEVQGGKMPKSYVKLISGVRKST 970
 77 QDIYPLTRHDELTLISGELITNIRKEDGKMWBGQJNGRRGLFPDVFBEIKKMKDP 136
 971 SIDTPTSPASLRVAPAKPAIPGEFI-----AMYTSESSQGLTF 1016
 137 LTNKAP-EKP-LHEV--PSQNSLSEETILRTNKGERRRRRCQVAFSLPQNDDEL 191
 1017 QCGDVI-VYTKQGDWMTGYGDKSGVFPSPNYV-----LKDSE 1054
 192 KVGDIIEVVGVEGVEGWGLNGKTGMFPSPNFKELSGESDELGISQDEQLSKSSLET 251
 1055 GSGTAGTSGSGKPELQAVIAYATGPBOL-----TLAPQGLIRKKNPGMWGE 1108
 252 GSESDGSS--STSGANGVATATAIQPKVKGVGCDLFXKPKILRRS-----IEVE 306
 1109 LQARKKQIQI-WFPANVYKLSFGTSKITPTELPTAVQPAVCQVIGMYDYRQNDDEL 1167
 307 NDPLFVEKTIKKLPAT---TATPDSSK--TEWDSRTSKDYCKVI--PPEKQNDDEL 358
 1168 AFSKQIINVLNKE--DPMWKGEVSGOVGLFPSPNYKL 1204
 359 TIKESDITVTLNKDCIDVGMWGBLNGRGVFPDNYKL 397

RESULT 15
 US-08-484-710-2
 Sequence 2, Application US/08484710
 Patent No. 5656438
 GENERAL INFORMATION:
 APPLICANT: Yen-Ming Hsu
 TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 State Street, Suite 510
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/484,710
 FILING DATE:
 CLASSIFICATION: 530

Fri Mar 14 14:49:32 2003

us-09-674-237a-3.rai

Page 11

```

?      ATTORNEY/AGENT INFORMATION:
?      NAME: Louis Myers
?      REGISTRATION NUMBER: 35,965
?      REFERENCE/DOCKET NUMBER: BGP-1900
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (617) 227-7400
?      TELEFAX: (617) 227-5941
?      INFORMATION FOR SEO ID NO: 2:
?      SEQUENCE CHARACTERISTICS:
?          LENGTH: 553 amino acids
?              type: amino acid
?          TOPOLOGY: linear
?              MOLECULE TYPE: peptide
?              FRAGMENT TYPE: internal
US-08-484-710-2

```

Query Match	5.3%	Score 333;	DB 1;	length 553;
Best Local Similarity	27.4%	Pred. No. 1.2e-15;		
Matches	93;	Conservative	67;	Mismatches 115;
			Indels	64;
			Gaps	15

```

QY 912 QATPAPAKKDNDHNFENKSDVITVLEBOD--MMMFEGVOQOKMFPKSYVKTLSGPRVST 970
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 77 QDIYPIYLRHDDDELITSVGEITITNIKKEDXWMEGQINGRGLFPDNEFRIKXMKKPD 136
QY 971 SIDGPTESPASLKRVA5PAKPAIPGEERI-----AMYVESSEOGDLTF 1016
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 137 LTNKAP--EKP--LHEV--PGNSLISEITLRTNKRGERRRRCQVAPSYLPQNDDEL 191
QY 1017 QQGDVIT-VYTKKDGDMWTGTVGDKSGVFPSPVYR-----LKDSE 1054
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 192 KVGDIIEVWGEVEGEGWEGVLNKGTFGPNFIKELSGEDELGISQDEQJLSKSLRFTT 251
QY 1055 GSGTAGKTGSLGKKPELAIQVIA5YATGPQL-----TLAPQLILFKKNPGSGWEGE 1108
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 252 GSEBDGDDSS--STSEBEGANCTVATAAIQPKVGVGFGDIIPDKPIKLPRSS---EIVE 306
QY 1109 LQAGKKRQIG-WEPANVYVLLSPGSKITPELPLRTAIVQPAVCQVIGMYDTAONDEL 1167
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 307 NDPLPVEKTIQKKLPAT--TATPDSSK--TEMDSRTSKDYCKVI--FPYEAONDEL 358
QY 1168 AFSKQIINVLNKE--DPMWKEG5GQVCLFPSNVYKL 1204
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 359 TIKEGDVIYTLINKDCIDVGMWBEELNGRRRCVFPDNEFVKL 397

```

```
Search completed: March 14, 2003, 12:18:27
Job time : 32 secs
```



TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELETYPE: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 193:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2873 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 US-08-630-915A-193

Query Match 6.6%; Score 333.2; DB 4; Length 2873;
 Best Local Similarity 61.0%; Pred. No. 1e-76;
 Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

QY 2867 CTTCCTGACCGTACCTAGTGTGCGCCAGTTACGGCAGAGATCAGCTTTACCCCAAGCA 2926
 Db 541 CTTTTCACAACTTAACCTTAATATCATCATGCGCAAGAAAATCAGCTTTCACCTGAACTG 600
 QY 2927 CAGCCACTGCTCTCTCCCATCTTCCCTCTCTGGCCGAGGGTGAAGGTGAGAGGCTAC 2986
 Db 601 TGTCCCTCTG---ATCTGTATCACTTAATCATGACAGGAGCAAGTGTAGAAAACCTTAA 657
 QY 2987 AAGGCAAGCCCTGTATCCCTGGAGAGCCAAAGAACACACTTAATTTTAAAGAAA 3046
 Db 658 AAGCAGAGCCCTTGTCTCTGAGCTCAAGAAAGAAATACACTTGAACCTTCTCAAAAC 717
 QY 3047 GTGACGTATCATCCGTTCTGTGAACAGCAAGACATGTGTGTTGAGAGAGTTCAAGTCT 3106
 Db 718 ATGACATTATTAATCTGTGTGAGAGCAGCAAGAAATGTGTGTTGAGAGAGTCAATGAG 777
 QY 3107 AGAAGGTTGTTTCCCAAGTCTTACTGAAACATTTTCAGGGCCCGCTAAGAAATCCA 3166
 Db 778 GAAGAGATGTTTCCCAATCTTATGTCAAGATCATTTCTGGAGATGAATTAAGAGGG 837
 QY 3167 CAACATGATGACTGCGCTTACTGAAAGTCTGTACTTAAGAGAGTGGCTTCCCGG 3226
 Db 838 AA-----GAACCAAGCTTTGATATGACGCTGAATAAAGAACTACCTCGG 885
 QY 3227 CCGCAGAGCCAGCATTCGCGAGAGAGTTTATTCGATGATACATACAGAGTTCTG 3286
 Db 886 CAGCTATTTAGT-----TGAAGAGATATATTTGCACTTTATCCATATTTCAAGTCTG 939
 QY 3287 AGCAAGAGATTTAACTTTAGCAAGAGGATGTGATTTGCTTACCAAGAGAGATGGT 3346
 Db 940 AACCTGAGATTTGACTTTCAAGAGAGTGAAGAAATATGTGTAAGCAAGAGATGGAG 999
 QY 3347 ACTGTTGACCGGAAAGCGTGGGCGCAAGTCCGAGTCTTCCCTTCTAATATGTGAGC 3406
 Db 1000 AGTGTGACAGAGATTTGAGATGAGATGAGATTTTTCATCAACATATGTCAAC 1059
 QY 3407 TTAAGATTCAGAGGCTCTGAACTGCTGAGAAAGAGGAGTTTGAAGAAAAAAGCTG 3466
 Db 1060 CAAGAGATCAAGAGATTTGGAGTCTGAGAGTCTGAGCATCAATTAAGAAAAAGCTG 1119
 QY 3467 AAATTTGCCAGATTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3526
 Db 1120 AGATTGCTCAGATTAATTTCAAGATATGTTGCTTCTGTTCTGAAACAACTTGAAGCT 1179
 QY 3527 CTGGGACAGCTATTTCTGATCCGAGAAAAAGAAAGAGTGTGATGTGAGAGAGAGAGCTG 3586
 Db 1180 CAGAGACATTAATTAATTTCTTAAGAAAAATACAGAGTGGGAGGAGAGAGAGATGATC 1239
 QY 3587 AAGCTGAG 3646
 Db 1240 AGGCGAG 1299
 QY 3647 GCGCCGAG 3706
 Db 1300 GTTCAAGATGATGAAG 1358

QY 3707 TGTCCAGGTGATCGGATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3766
 Db 1339 TATGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
 QY 3767 GCAAGGCGCATGATCAAGTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3826
 Db 1399 CCAAGGAGCACTATTAATTTATGAAAGATGATCTGATGATGATGATGATGATGATGATGAT 1458
 QY 3827 TCAGTGGCAAGTTGGGCTTCTTCCATTCATTAATGTTAAAGTGTGACAGAGAGAGAGAG 3886
 Db 1459 TCAAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1518
 QY 3887 CCAGCCAGCAATGATCA 3904
 Db 1519 CAAGTCAAGTGAAG 1536

RESULT 3
 US-08-630-915A-39
 Sequence 39, Application US/08630915A
 Patent No. 6309820

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
 APPLICANT: HOFFMAN, No. 6309820h
 APPLICANT: KAY, Brian K.
 APPLICANT: FOWLES, Dana M.
 APPLICANT: McCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 NUMBER OF INVENTIONS: 227
 CORRESPONDENCE ADDRESS:
 ADDRESS: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/630, 915A
 FILING DATE: 03-APR-1996
 CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:
 LENGTH: 747 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-630-915A-39

Query Match 5.1%; Score 258; DB 4; Length 747;
 Best Local Similarity 63.9%; Pred. No. 1.8e-57;
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

QY 3247 GGAAGAGATTTATTTCCATGATACATACGAGATTTCTGAGCAAGAGATTTAACTTT 3306
 Db 115 GGAAGAGATTTATTTGACATTTATTCATATTCAGAGTGTGAAGCTGAGATTTGACTTTC 174
 QY 3307 CAGCAAGGATGTGATTTGTTTACCAAGAAAGATGTGAGTGTGAGAGAGAGAGAGAGAG 3366

```

Db 175 ACAGAGGTGAAGAAATATTTGTGACCCAGAAAGATGAGAGTGTGAGACAGAAATATT 234
QY 3367 GGCACAAAGTCCGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGTTGAGAGGCTCT 3426
Db 235 GGAATATGAAGTGAATTTTTCATCAAACTATGTCAAAACCAAGATCAAGAGAGTTT 294
QY 3427 GGAATCTGTGGAAACAGGAGTTTAGAAAAAACCTGAAATTTGCCAGTTATTTGCT 3486
Db 295 GGGAGTGTGACAGTCTGGAGCATCAATTAATAAAACCTGAGATTTGCTCAGTAACTTCA 354
QY 3487 TCCAGCTGTACTGTGTCGCCAACAACCTACCTGGCTCTGGGAGCTGATTCGATC 3546
Db 355 GCATATGTTGTTCTGTGTTCTGAACAATTAGCCTTGACAGACAGTTAAATTAAT 414
QY 3547 CGGAAAAAAGCCAGGTGATGTTGGAAAGAGAACTGCAAGTCAAGGAAAAAGCC 3606
Db 415 CTAAAGAAAAATACAGTGGGTGGTGGCAAGAGATTACAGGCCAGAGSAAAAAGCCA 474
QY 3607 CAGATAGGTTGTTTCCAGCAAAATATGTCAAACTTCTAAGCCCGGAACAGCAAAATC 3666
Db 475 CAGAAAGATGTTTCTGTCAGTCATGTTAACTTTGGTCCAAGCATGMAAGAGCC 534
QY 3667 ACCCACTAGCTACCCAGACCGCAGTGCAGCAGCATGTCAGGATCGGGATG 3726
Db 535 ACACCTGCTTTCATCTCT-----GTATGTACAGTGAATTCCTATG 573
QY 3727 TAGATTAACACCCCGCAAGACATGACGAATAGCTTCAGCAAAAGCCAGATCATCAAC 3786
Db 574 TAGTACTATGACCAATATATGAAGATGAGCTCAGTTTCTCAAGGACCACTCATTAAT 633
QY 3787 GTCTCTCAAGAGGAGCCCGGAGTGTGGAAGAGAAATCGATGGGCAAGTTGGCTTC 3846
Db 634 GTTATGAACAAAGATGATCTCTGATTTGGTGCAGAGAGATCAACGGGTGACTGTCTC 693
QY 3847 TTCCCATTCATTTATGTAAGCTGACACAGACATGAGACCCCAAGCAATGA 3900
Db 694 TTTCCTTCAACTAGCTTAGATGACAGACGACTCAGATCCAAAGTCAACAGTGA 747

RESULT 4
US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

Query Match 3.0%; Score 151.4; DB 4; Length 531;
Best Local Similarity 61.3%; Pred. No. 9.5e-30;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

```

```

QY 226 GGGGTGATCAGCAAGTAACTATATACACATGCTCACTTTCCACACTTTTGGT 285
Db 9 GGCTGAGCTGCAAGAAAGTCAAGATCATGATGGCTCACTTTCCCAACGATGAAT 68
QY 286 GGATGCTGATGTCTGGCCATAACTGTGAGAGAAAGGCGCAAGCATACAGAGTTC 345
Db 69 GGAGGGCCAAATATGTGGCTATTATCTGAAGAAAGTAAAGCATATATAAAGATT 128

```

```

QY 346 CTTAGCTGAAGCCGATAGCCGGAATTTATCTGTGATCAAGGAGAACTTTTTC 405
Db 129 GATAACCTCAAACTTCAAGAGGTTTACATTAACAGTATACAGCCCGTACTTTTCTTA 188
QY 406 CAATGTGGTTACTCTACGCTGCTTACACAAATATGGCGCTTAGGAGCATGATAC 465
Db 189 CAGTAGGCTGTGCGGCCCGCTTTTAGCTGAATATAGGGCTTATCATCTTGAACAG 248
QY 466 GATGAGAGATGATCAAGTGAATTTTCCATAGCCATGAAGCTTATCAAGTAACTGA 525
Db 249 GATGGAAGATGACACAGCAAGAGTTCTTATAGCTATGAATCATCATTAAGTTG 308
QY 526 CAAGATATACGTTCCCTCCACATTCCTCCCTGTCATGAAACAGCAACAGTGTATT 585
Db 309 CAGGGCCAAAGCTGCTGTGATGCTCTCCCTCTATCATGAACAAACCC-----CTATG 362
QY 586 TCCAGTCAACGAGATTTTGTATAGAGGATTTCTAGATCCACATCAGACTGT 645
Db 363 TTCTCTCACTAATCTCTGCTGTTTGGATGGAGGATGCCCAATGTGTCATTCAT 422
QY 646 GCTCCTGTGCC 656
Db 423 CAGCATTTGCC 433

RESULT 5
US-09-404-879A-60
; Sequence 60, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-60

Query Match 2.8%; Score 144; DB 4; Length 480;
Best Local Similarity 62.1%; Pred. No. 7.6e-28;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```

```

QY 259 ATGGCTAGTTTCCCACTCTTGGGTGAGCTGTGATGTCTGGGCATTAATCTTGAG 318
Db 14 ATGGCTAGTTTCCCAAGGATGAATGAGGCGCAAAATATGTGGCTATTATCTGAA 73
QY 319 GAAAGGCGCAAGATGACCGAGTCTCTTAAGCTGAAGCCGATAGCGGATTTATTA 378
Db 74 GAAGTACTAAGCATGATTAACAGTTTGATTAACCTTAACCTTCAAGAGTTAGATA 133
QY 379 GGTATGAAGGAGAACTTTTTCATCTGGGTACTTCAAGCTGTCTTGAACAA 438
Db 134 GGTATGAAGCCCGTACTTTTTCCTACAGTCAAGTCTGCGGCCCGGTTTATGCTGA 193
QY 439 ATATGGCGCTAGCGGACATGATTAAGATGAAGATGATTAAGTAAATTTTCCATA 498
Db 194 ATATGGCTTATCAATCTGAAACAGAGATGGAGATGACAGCAAGATTTCTTATA 253
QY 499 GCGATGAAGTTTCAAACTGAAGCTTAAAGATATACAGTCCCTCCACACTTCCCT 558
Db 254 GCTATGAAGTCAATCAAGTTAAAGTTGACAGGCGCAACAGCTGCTGTATCTCTCT 313
QY 559 GTATGAAAGACCAACAGTGGCTATTTCCATGTCACGACCATTTGTATGAGAGATT 618
Db 314 ATCATGAACCAACCC-----CTATGTTCTCTCACTAATCTCTGCTGTTTGGATG 367

```

QY 619 GCTAGCATGCCACCACTCAAGCTGTGCTCTGTC 656
Db 368 GGAAGCATGCCCAATCTGTCATTCAATCAAGCCATTGCC 405

RESULT 6
US-08-728-323A-1

; Sequence 1, Application US/08728323A
; Patent No. 5948676

GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A

FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSK/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489

US-08-728-323A-1

Query Match 2.8%; Score 144; DB 2; Length 3489;
Best Local Similarity 45.1%; Pred. No. 2.5e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 1233 GGATCAGAGGCTGCTTGAAGAGCCGCTGTCAGAGATGAGCAGCAGCAGAGAGAAACT 1292
Db 1578 GCAAGAGCCACAGCAGAGAGCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1637
QY 1293 GCGCTGACATTTGAATATAAGAGCGGGAAGACTTGAAGCAGCAGCAGCTTGAAGCTGGA 1352
Db 1638 GCAGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCAGCAGCGGAGCCACAGCA 1697
QY 1353 GAAGCGCCGCAAGCGCTTGTGAGCAGCAGCGCAAGAGCAGAGAGCGTTGCTCAGCT 1412
Db 1698 GCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757
QY 1413 GGAAGCGCCGAGAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGCGCAAGCGCGCA 1472

Db 1758 GCGGAGCCACAGCAGCGGAGCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1817
QY 1473 GCTGAGCTGGAAGAGCAGCTGGAAGAGCAGCGGAGCTGAGCGGAGCAGAGAGAGGA 1532
Db 1818 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1877
QY 1533 GAGGAGAGAGAGATCGAGAGCGCCGAGCCGCAAAACGGAACTGGAAGAGCGCACA 1592
Db 1878 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1937
QY 1593 ACTGAATGGAAG 1652
Db 1938 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1997
QY 1653 GGGCAGCTGCTGCTGGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712
Db 1998 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2057
QY 1713 TGACAAAAGCAGCAGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1772
Db 2058 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2117
QY 1773 GAGGAG 1832
Db 2118 GAGGAG 2177
QY 1833 CTACAGCAGCAGGTTGAG 1892
Db 2178 GAGGAG 2237
QY 1893 GATGACTGATGAG 1952
Db 2238 GCAGCAGCAGCAGAGATGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2297
QY 1953 TACCTCAAAAGAGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2012
Db 2298 GCAGGAG 2357
QY 2013 GAGCAGGAG 2072
Db 2358 GAGGAG 2417
QY 2073 GCTGAAG 2132
Db 2418 GCAGGAG 2477
QY 2133 GGCAGCGGAGCTGAAG 2192
Db 2478 GAGGAG 2537
QY 2193 GGAAGAGCCTCAG 2252
Db 2538 GCAAG 2597
QY 2253 GAGGAG 2312
Db 2598 GAGGAG 2657
QY 2313 TGTCAAG 2372
Db 2658 GCAGGAG 2717
QY 2373 GCTTTTCATCCGATCAG 2416
Db 2718 GGAAG 2761

RESULT 7
US-09-298-568-1

; Sequence 1, Application US/09298568
; Patent No. 6322792

GENERAL INFORMATION:

APPLICANT: Kieff, Elliott D.

```

; APPLICANT: Bailestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

```

```

Query Match      2.8%; Score 144; DB 4; Length 3489;
Best Local Similarity 45.1%; Pired. No. 2.5e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

```

```

QY 1233 GGATTCAGAGGCTGCTTGAAGAGCGCGTCTGAGAGATGAGCAGCAGCAGAGAAAGT 1292
DB 1578 GCAGGAGCCACAGCAGCAGAGAGCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1637
QY 1293 GCCTGTGACATTTGAAGATTAAGAGCGGAGAACTTGAAGCAGAGCAGTGTGAGCTGA 1352
DB 1638 GCAGGAGCCACAGCAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGA 1697
QY 1353 GAAGCGCGCCAGCAGCGCTCTTGAAGCAGCGCCGAAAGCAGCAGCAGCGTGTGCTAGCT 1412
DB 1698 GCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGA 1757
QY 1413 GAGAGCGCGCCAGCAGCAGAGAGAAAGCGGAGCGCCAGCAGCAGAGCCAAAGCGCA 1472
DB 1758 GCGGAGCCACAGCAGCGGAGCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1817
QY 1473 GCTTGAGCTGGAAGAGCAGCTTGAAGAAAGCAGCGGAGCTGAGCGGAGCCAGAGAGGA 1532
DB 1818 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCA 1877
QY 1533 GAGGAGGAAGAGATGAGAGGCGCGGAGCGCCGAAACGGGAAGCTGAAAGAGCAGCA 1592
DB 1878 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCA 1937
QY 1593 ACTTGAATGGAAAGAGAGCGAGAGCAGAACTCTGAATCAGAGAGAAAGAGCAGCA 1652
DB 1938 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 1997
QY 1653 GGGCAGCTGTGCTCTGAAGCAGAGAGAGAGCTTGAAGTTGAGTTAGAGCTTGA 1712
DB 1998 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 2057
QY 1713 TGCAAAAAGCATCAGCTAGAGAGAAAGCTTCAAGATATAGCTGTGCACTGCAACCA 1772
DB 2058 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2117
QY 1773 GAGGCAAGAAATGAGAGCAGAAAGTCTTGAAGAGCTTGAAGTTGCTGAATCAACCA 1832
DB 2118 GAGGCAAGAGATGAGCAGCAGAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCA 2177
QY 1833 CTTAAGCAGCAGATTTGAGAGATCTCAGCAATGCTTGAAGAGCTTATTCAGAGAAACA 1892
DB 2178 GATGAGCAGCAGAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2237
QY 1893 GATAGCAGCAGCAGTAAAGCAAGTCCAGCAGAAAGTTCATAGAGACTCGTTCT 1952
DB 2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGAGAGCAGCAGCAGCAGAGAGAGAGCA 2297
QY 1953 TACCCTAAAAAGAGCTTGAAGAGAAAGAGCTGGCCCGCAGCAGCTCCGGAGCAGCT 2012
DB 2298 GCAGGATTTGAGAGAGCAGAGAGAGATTAGAGATCAGAGCAGAGATTAGAGAGAGCA 2357

```

```

QY 2013 GGACGAGGTGAGAGAGAGCAGCTCAAGCTGACAGAGATTGATTTTCAACACCA 2072
DB 2358 GAGAGAGAGATTAGAGAGCAGCAGAGAGATTAGAGAGCAGCAGAGAGATTAGAGCA 2417
QY 2073 GCTGAAGGAATGAGAGATACCTAGCAACCACTCCAGAGCAGAGCTCCTGGA 2132
DB 2418 GCAGGAGCAGAGATTAGAGAGCAGCAGAGAGATTAGAGAGCAGAGAGATTAGCA 2477
QY 2133 GGCAGCGGCACTGAAGCAGAAAGAGCAGAGAGAGAGAGAGCTGAGTTAGAGAGCAAAA 2192
DB 2478 GAGCAGAGCAGAGATTAGAGAGCAGAGAGAGAGAGAGATTAGAGAGCAGAGAGAG 2537
QY 2193 GGAAGAGCCTCAGAGCAGAGATTGAGAAAGAGCAGCAAGCAATGCTGAGAGATGAGCA 2252
DB 2538 GCAAGAGCAGAGAGTGAAGAGCAAGAGCAGAGCAGAGAGAGAGAGAGATTAGAGAGGT 2597
QY 2253 GAGAGAGCAGCAGCGCCCGGAAACCCACAGAGAGAGCAGACTGAAGAGGAGAGCAG 2312
DB 2598 GAGAGAGCAGAGAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGATTAGAGAGGTGAAGA 2657
QY 2313 TGTCAGAAAGAGAGCGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGC 2372
DB 2658 GAGAGAGAGCAGAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2717
QY 2373 GCTTTCCATCCGATAGAGAGCAGCTTAAGCTGCGCAGCAGG 2416
DB 2718 GGAAGAGCAGAGCAGCAGAGAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGG 2761

```

RESULT 8
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564

GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 2.8%; Score 144; DB 2; Length 32207;
 Best Local Similarity 45.1%; Pred. No. 9.8e-27;
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 1233 GATCAGAGGCTCTGAGAGCCGCTGTCAGAGATGAGCAGCAGCAGAGAGAACT 1292
 DB 20419 GCAGAGGCCACAGCAGAGAGCCACAGCAGAGCCACAGCAGAGAGCCACAGCA 20360
 QY 1293 GCTGTGACATTTGAAGATTAAGACCGGAGAACTTCAGCCGAGGAGTGTGAGCTTGA 1352
 DB 20359 GCAGAGGCCACAGCAGCGGAGCCAGCAGCGGAGCCAGCAGCGGAGCCAGCA 20300
 QY 1353 GAAGCCGCCGCAAGCCGCTTTGAGCAGCAGCCGCAAGAGCAGAGAGCTGCTCAGCT 1412
 DB 20299 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 20240
 QY 1413 GAGCGCGCCGAGCAGAGAGAGAAAGCCGAGCCGAGCAGCAGAGAGAGCCAGCGCA 1472
 DB 20239 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 20180
 QY 1473 GCTGTGAGCTGGAAGAGCAGCTGAGAAAGCAGCGGAGCTGAGAGCCGAGCAGAGAGCA 1532
 DB 20179 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 20120
 QY 1533 GAGGAGGAAGAGATGAGAGGCGCGGAGCCGCAAGAGCTGAAAGAGCAGCAGCA 1592
 DB 20119 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 20060
 QY 1593 ACTTGAATGGGAAACGAAACCGAGACAGAACTCTGAAATCAGAGAAACAGAGAGCA 1652
 DB 20059 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 20000
 QY 1653 GGGCAGCGTGTCTTGAAGGCAAGAGAGAACTCTGAGATTTGAGTTGAAGCTTGA 1712
 DB 19999 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 19940
 QY 1713 TGACAAAAGCATCAGTAAAGAAATCTCAGATATCAGTGTGCACTGCAACCA 1772
 DB 19939 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 19880
 QY 1773 GAGGCAAGAAATTTAGAGACGAACTGTAGAGCTTAAGATTCTGAAATCAACCA 1832
 DB 19879 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 19820
 QY 1833 CTTAAGCAGCAGTGTGAGAGATCTCAGCAATGCTTGAAGACTTATTTCAAGAAACA 1892
 DB 19819 GATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 19760
 QY 1893 GATGAGCAGCAGCAGTAAACAAAGTCCAGCAAGCACTTTCAGTGAAGCTCGCTTCT 1952
 DB 19759 GCAGAGTGAACACAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 19700
 QY 1953 TACCTCAAAAGAGCTTTGAAGCAAGAGAGCTGCGGAGCAGAGCTCCGGAGAGAGCT 2012
 DB 19699 GCAGAGTGAAG 19640
 QY 2013 GAGCAGGTGAG 2072
 DB 19639 GAGCAG 19580
 QY 2073 GCTGAAGAACTGAGAGATACATACCAACAGCACTCCAGAGAGAGAGAGAGAGAGAG 2132
 DB 19579 GCAGAGTGAAG 19520
 QY 2133 GGCAGCCGCACTGAAG 2192
 DB 19519 GGCAG 19460
 QY 2193 GGAAGAGCCTTGAAG 2252
 DB 19459 GCAAG 19400

QY 2253 GAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2312
 DB 19399 GAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19340
 QY 2313 TGTCAAGAAAG 2372
 DB 19339 GCAG 19280
 QY 2373 GCTTTTCATCCGATCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2416
 DB 19279 GGAAGAGCAG 19236

RESULT 9

US-08-757-669A-20/c
 ; Sequence 20, Application US/08757669A
 ; Patent No. 6183751
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/757,669A
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 45185-F
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 32207 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-757-669A-20

Query Match 2.8%; Score 144; DB 4; Length 32207;
 Best Local Similarity 45.1%; Pred. No. 9.8e-27;
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 1233 GATCAGAGGCTCTGAGAGCCGCTGTCAGAGATGAGCAGCAGCAGAGAGAACT 1292
 DB 20419 GCAGAGGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 20360
 QY 1293 GCTGTGACATTTGAAGATTAAGACCGGAGAACTTCAGAGAGAGCAGTGTGAGCTTGA 1352
 DB 20359 GCAGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGGCCAGCAGCGGAGCCACAGCA 20300
 QY 1353 GAGCGCGCCGAGCAGCGCTCTTGAAGCAGCAGCAGCAGCAGCAGAGAGAGAGAGAG 1412
 DB 20299 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 20240

```

QY 1413 GGAGCGCGCCGAGCAGAGAGAGAAAGCGGAGCCCGCAGAGCAGAGGCGCAAGCGCA 1472
Db 20239 GCGGAGAGCCACAGACCGGAGGCCACAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA
QY 1473 GCTTGAGCTTGAGAGCAGCTTGAGAGACCGGAGCTTGAGCGGCGAGCGAGAGAGA 1532
Db 20179 GCAGGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA
QY 1533 GAGAGAGAGAGATGAGAGAGCGGAGCGCCGAGAAACCGGAACTTGAGAGAGCAGAGCA 1592
Db 20119 GCAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCA
QY 1593 ACTTGATGAGAAACGGAACCGAGACAGGAACCTCTGAATCAGAGAGAAACAGAGAGAGA 1652
Db 20059 GCAGGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGAGATGAGCAGCA
QY 1653 GGGCAGCCGTGCTCTTGAGAGCAAGAGAACTCTTGAGATTGAGTTAGAACTCTGAA 1712
Db 19999 GCAGGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA
QY 1713 TGCAAAAAGCATCAGCTGAAAGAAAACCTTCAGGATATCAGGTGTCGACTGGCAACCCA 1772
Db 19939 GCAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGCAGAGATGAGCA
QY 1773 GAGGCAAGAAATTTGAGAGCAGCAAGAGCTTGAAGAGCTTGAATTTGCTGAATCACCAC 1832
Db 19879 GAGAGCAGCAGATGAGAGAGAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCA
QY 1833 CTTTACAGCAGCAGATTTGAGAGATCTCAGCAAACTGCTTGAGAGACTTATTCAGAGAAA 1892
Db 19819 GAGTGAAGCAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGAGATGA
QY 1893 GATACCTAGTGAACAGTTAAACAACTCCAGCAGCAAGTTTCATAGAGACTCGCTTCT 1952
Db 19759 GCAGCAGCAGCAGATGAGAGAGAGCAGCAGAGAGCAGCAGCAGCAGAGAGAGAGCA
QY 1953 TACCTCAAAAAGACCTTGAGAGCAAAAGAGCTGGCCCGGAGCAGCTCGGAGAGCAGCT 2012
Db 19699 GCAGGATTTAGAGAGCAGAGAGCAGAGATTAGAGATCAGAGCAGAGATTAGAGAGAGCA
QY 2013 GAGCAGAGTGAAGAGAGAGCCAGCTCAAAAGCTGCGAGAGATTGATTTTCAACAGCA 2072
Db 19639 GAGCAGAGATTTAGAGAGCAGAGAGAGAGATTAGAGAGCAGAGCAGAGATTAGAGAG
QY 2073 GCTGAAGAACTGAGAGAGATATCATGCAAAACAGCAACTCCAGAAACAGAGTCCCTGGA 2132
Db 19579 GCAGAGCAGAGATTAGAGAGAGAGAGAGAGAGATTAGAGAGCAGAGAGAGATTAGA
QY 2133 GGCAGCGCGAGCTGAAGCAGAAAGAGCAGAGAGAGAGAGAGCTTGAGTTAGAGAGCAAAA 2192
Db 19519 GAGAGCAGAGAGATTAGAGAGAGAGAGAGAGAGAGATTAGAGAGCAGAGAGTGGAGAA
QY 2193 GGAAGAGCGCTCAGAGAGAGATTAGAGAAAGGAGCAAGCAGATGGCTGAGCATGTGAGCA 2252
Db 19459 GCAAGAGCAGAGAGTGAAGAGCAAGAGCAGAGAGAGAGAGAGAGAGATTAGAGAGAGT
QY 2253 GGAAGAGCAGCAGCGCCCGGAGAAACCCACAGAGAGAGCAGAGCTGAAGAGAGAGAGAG 2312
Db 19399 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTAGAGAGAGTGGAGAA
QY 2313 TGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGC 2372
Db 19339 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGC
QY 2373 GCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCAGCCAGG 2416
Db 19279 GGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19236

```

```

; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bozenky, Roy A
; APPLICANT: Russo, James U
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

```

```

Query Match          2.8%; Score 144; DB 4; Length 32207;
Best Local Similarity 45.1%; Pred. No. 9,8e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

```

```

QY 1233 GGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGATGAGCAGCAGCAGCAGAGAAACT 1292
Db 20419 GCAGAGGCCACAGAGAGAGAGAGCCACAGCAGAGAGCCACACAGCAGAGAGAGAGCA
QY 1293 GCTTGAACATTTGAGAGATGAAGAGCGGAGAACTTCAGAGCAGAGAGAGTGAAGTGA 1352
Db 20359 GCAGAGGCCACAGAGAGCGGAGGAGCAGCAGAGCGGAGCGCCACAGCAGCGGAGCAGCA
QY 1353 GAAAGCGCGCGCAAGCGCTCTTGAGAGCAGCAGCGCAAGAGCAGAGAGCGGTGCTCAGCT 1412
Db 20299 GCGGAGCAGCAGAGAGGAGAGAGCCACAGCAGCGGAGAGCCACAGCAGAGGAGCAGCA
QY 1413 GGAAGCGCGCGAGCAGAGAGAGAGAAAGAGCGGAGAGCGCCAGAGCAGAGAGCCAGAGCGCA 1472
Db 20239 GCGGAGCAGCAGAGAGGAGAGAGCCACAGCAGCGGAGAGCCACAGCAGAGGAGCAGCA
QY 20239 GCGGAGCAGCAGAGAGCGGAGGAGCCACAGCAGAGATGAGCAGCAGAGAGATGAGCAGCA 20180
QY 1473 GCTGAGCTGAGAGAGAGAGCTGAGAGAGAGCGGAGAGCTTGAGAGCGGAGAGAGAGAGAG 1532
Db 20179 GCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGAGCA
QY 1533 GAGAGAGAGAGATGAGAGAGCGGAGCGCGAGAGCGCAAAACGAGAACTGAGAAAGCAGCAGCA 1592
Db 20119 GCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGAGCA
QY 1593 ACTTGAATGGAGAACGAAACCGAGAGCAGAACTCTGAATCAGAGAGAAACAGAGAGAGAG 1652
Db 20059 GCAGAGATGAGAGAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGAGAGCA
QY 1653 GGGCAGCGTGTCTTGAAGGCAAGAGAGAGAGAGCTTGAGATTGAGATTGAGAACTTGAA 1712
Db 19999 GCAGAGATGAGAGAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGAGCA
QY 1713 TGCAAAAAGCATCAGCTGAAAGAAAACCTTCAGGATATCAGGTGTCGACTGGCAACCCA 1772
Db 19939 GCAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGCAGAGATGAGCA
QY 1773 GAGGCAAGAAATTTGAGAGCAGCAAGAGCTTGAAGAGCTTGAATTTGCTGAATCACCAC 1832
Db 19879 GAGAGCAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGC
QY 1833 CTTTACAGCAGCAGATTTGAGAGAGATCTCAGCAAACTGCTTGAGAGACTTATTCAGAGAAA 1892
Db 19819 GAGTGAAGCAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGAGAGAGAGTGC
QY 1893 GATACCTAGTGAACAGTTAAACAACTCCAGAGAAACAGTTTGCATAGAGAGACTCGCTTCT 1952
Db 19759 GCAGCAGCAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19700

```


QY 1953 TACCTCAAAAGAGCCTTGAAAGCAAGAGCTGCCCCGAGCAGCTCCGGAGCAGCT 2012
 DB 19699 GCAGAGATTAGAGAGCAGAGAGAGATTAGAGGATCAGAGCAGAGATTAGAGAGCA 19640
 QY 2013 GCAGAGCTGAGAGAGAGAACCAAGCTCAAAAGCTCCAGAGATTGATGTTTCAACACCA 2072
 DB 19639 GGAGCAGGAGTTAGAGAGCAGAGAGAGATTAGAGGAGAGCAGAGATTAGAGGA 19580
 QY 2073 GCTGAAGGAAGCTGAGAGATCATAGCAAAACAGCAATCCAGAGCAGAGTCCCTGGA 2132
 DB 19579 GCAGAGCAGAGATTAGAGAGAGCAGAGAGATTAGAGGAGCAGAGAGATTAGA 19520
 QY 2133 GCGAGCGCAGCTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2192
 DB 19519 GCGAGCAG 19460
 QY 2193 GGAAGAGCTCAG 2252
 DB 19459 GCAAGAGCAG 19400
 QY 2253 GGAAGAGCAG 2312
 DB 19399 GGAAGAGCAG 19340
 QY 2313 TGTCAAG 2372
 DB 19339 GCAG 19280
 QY 2373 GCTTTTCATCCGATCAG 2416
 DB 19279 GGAAGAGCAG 19236

RESULT 11

US-09-404-879A-4/C
 ; Sequence 4, Application US/09404879A
 ; Patent No. 6468546
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Algate, Paul A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
 ; FILE REFERENCE: 210121.462C2
 ; CURRENT APPLICATION NUMBER: US/09/404,879A
 ; CURRENT FILING DATE: 1999-09-24
 ; NUMBER OF SEQ ID NOS: 393
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 531
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(531)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-404-879A-4

Query Match 2.4%; Score 121.6; DB 4; Length 531;
 Best Local Similarity 53.5%; Pred. No. 5.6e-22;
 Matches 250; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 1642 AAG 1701
 DB 531 AAG 472
 QY 1702 GAAGCTCTGATGACAAAAGCATCAGTAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
 DB 471 GAAGCACTGATGAGAAAGCATCAGCAATCTCAGGAGAGAGAGAGAGAGAGAGAGAGAG 412
 QY 1762 CTGGCAACCCAG 1821

DB 411 AAGCAAAACAAAAGAGCTGAGCTAGAGAGTTTGGATTAACAGTGTACCTGAAATTATG 352
 QY 1822 GAATATCCCACTTACAGCAGCAGAGTTGACAGAAATTCAGCAAAATGCTTGAGAGATTAT 1881
 DB 351 GAATCAAAACATTTCAACAAAGAGCTTAAGAAATATCAAAATTAAGCTTATCTGTCGTC 292
 QY 1882 CCAGAGAAACAGATCTCAGTACAGCACTTAAACAAAGTCCAGAGAGAGAGAGAGAG 1941
 DB 291 CTTGAGAGAGAGCTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
 QY 1942 GACTGCTTTACCTCAAAAGAGCTTGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAG 2001
 DB 231 TCAGAGATCAGTTACTTACTTCAATAAAGATCATCAGAAAGAGAGAGAGAGAGAGAG 172
 QY 2002 GCGAGAGAGCTGAG 2061
 DB 171 AAGAGCAATTAATGATGCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112
 QY 2062 TTCAACAACAGCTGAAG 2108
 DB 111 TTTAACAATCAGCTGAAG 65

RESULT 12

US-08-056-200-93
 ; Sequence 93, Application US/08056200
 ; Patent No. 5616500
 ; GENERAL INFORMATION:
 ; APPLICANT: Steinert, Peter M.
 ; APPLICANT: Lee, Seung-Chul
 ; APPLICANT: Kim, In-Gyu
 ; APPLICANT: Chung, Soo-Il
 ; APPLICANT: Park, Sang-Chul
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
 ; TITLE OF INVENTION: Methods of Using Same
 ; NUMBER OF SEQUENCES: 117
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive, Sixteenth Floor
 ; CITY: Newport Beach
 ; STATE: CA
 ; COUNTRY: U.S.A.
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/056,200
 ; FILING DATE: 30-APR-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fedrick, Michael F.
 ; REGISTRATION NUMBER: 36,799
 ; REFERENCE/DOCKET NUMBER: NIH054.001A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (714) 760-0404
 ; TELEFAX: (714) 760-9502
 ; INFORMATION FOR SEQ ID NO: 93:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9551 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1507..1644
 ; FEATURE:
 ; NAME/KEY: intron

LOCATION: 1645..2511
FEATURE: CDS
NAME/KEY: 2512..8070
LOCATION: 2512..8070
US-08-056-200-93

Query Match 2.4%; Score 119.6; DB 1; Length 9551;
Best Local Similarity 44.1%; Pred. No. 1.1e-20;
Matches 500; Conservative 0; Mismatches 634; Indels 0; Gaps 0;

1233 GCATCAGAGCTGCTTGAAGAGCCCTGCTGATGAGATGAGCAGCAGCAGAGAAACT 1292
3324 GCAG 3383
1293 GCTGTGACATTTGAAATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1352
3384 GCAGCTAG 3443
1353 GAAGCCGCGCAAG 1412
3444 GAG 3503
1413 GAG 1472
3504 GAG 3563
1473 GCTGAG 1532
3564 GCGGAG 3623
1533 GAG 1592
3624 GCAG 3683
1593 ACTTGAATGAG 1652
3684 GCTGAG 3743
1653 GGGGAG 1712
3744 CGAG 3803
1713 TGACAAAG 1772
3804 GAAGCCGAG 3863
1773 GAGGAG 1832
3864 CGAG 3923
1833 CTTAG 1892
3924 GCAG 3983
1893 GATAG 1952
3984 GCTGAG 4043
1953 TACCCTTAAAG 2012
4044 AAG 4103
2013 GAG 2072
4104 GCAG 4163
2073 GCTGAG 2132
4164 GCTGAG 4223
2133 GGCAG 2192
4224 CGAG 4283

2193 GGAAGAGCTCAG 2252
4284 GCGGTGAAGCGCAG 4343
2253 GAG 2312
4344 GCAG 4403
2313 TGTCAG 2366
4404 GCTGAG 4457

RESULT 13

US-08-800-644-93
Sequence 93, Application US/0800644
Patent No. 5958752
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,644
FILING DATE: 14-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/056,200
FILING DATE: 30-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-800-644-93

Query Match 2.4%; Score 119.6; DB 2; Length 9551;
 Best Local Similarity 44.1%; Pred. No. 1.1e-20;
 Matches 500; Conservative 0; Mismatches 634; Indels 0; Gaps 0;

1233 GATTCAGAGCTGCTGAGAGCGCTTCCTACAGAGTGTGACGACCCAGAGAAACT 1292
 3324 GCAGGAGGAGGCGGACAGAGAGGCGCGAGCAGAGAGAGAGCGCGAGCA 3383

1293 GCTGTGACATTTGAAGATTAAGAAAGCGGAGAACTTCAGCAGGCGAGTGTGAGCTGGA 1352
 3384 GCAGCTGAGAGCGCGAGAGAGAGAGGCGCGAGCAGCAGCTGAGCGCGAGAGAGGA 3443

1353 GAAGCGCGCGCAAGCGCTCTTGAGCAGCAGCGCAAAAGACAGAGCGGTTGCTCAGCT 1412
 3444 GAGAGAGCGCGAGCAGCAGCTGAGCGCGAGCAGAGAGAGAGCGCGAGCAGAGCT 3503

1413 GAGAGCGCGCGAGCAGAGAGAGAGAGCGGAGCGCGCAGAGCAGAGAGAGAGAGAG 1472
 3504 GAGCGCGCGAGCAGAGAGAGAGAGCGCGAGCAGCAGCTGAGCGCGAGCAGCAGCTGAG 3563

1473 GCTGAGAGCTGAGAGAGCAGCTGAGAGAGCGGAGCTGAGCGCGCAGAGAGAGGA 1532
 3564 GCAGGAGCAGCAGCTAGAGCGCGAGCAGCAGCTGAGCGCGCAGCAGCAGCTGAGCGCGCA 3623

1533 GAGAGAGAGAGAGATCAGAGAGCGCGAGCGCGCAAAACGAGAACTTGAAAGCAGCGCA 1592
 3624 GCAGCAGCTGAGAGCGCGAGAGCAGCTGAGCGCGAGCAGCAGCTGAGCGCGAGAGCA 3683

1593 ACTTGAAATGGAG 1652
 3684 GCTGAGAGCGCGAGCAGCAGCTGAGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 3743

1653 GGGCAGCGTGTCTCTGAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712
 3744 CGAGCAGAGAGAGCGCGAGCAGCGCTGAGAGCGCGAGCAGAGAGAGAGAGCGGATGTGCT 3803

1713 TGAAG 1772
 3804 GAGAGCGCGAG 3863

1773 GAGGAG 1832
 3864 CGAGCAG 3923

1833 CTAGCAGCAGCAGTGTGAG 1892
 3924 GCAGCAGAGAGAGCGCGAGCAGCACTTAAGCGCGAGCAAGAGAGAGAGAGCGCGAGAGCG 3983

1893 GATACCTCAGTGACCAAGTTAAACAAAGTCCAGCAGAAAGCTTTGATAGAGAGCTCGCTCT 1952
 3984 GCTGAAGCGCGAG 4043

1953 TACCTTCAAAAGAGCCTTGAGAGCAAAAGAGCTGCGCGCGAGCAGCTCCGGAGAGAGCT 2012
 4044 AAGACCGAGCAG 4103

2013 GAGCAGAGTGAAG 2072
 4104 CGAGCAG 4163

2073 GCTGAAG 2132
 4164 GCTGAAGCGCGAG 4223

2133 GAGAGCGCGAG 2192
 4224 CGAGCAGCGAG 4283

2193 GAG 2252
 4284 GCGGCTGAAGCGCGAG 4343

2253 GAG 2312

4344 GCAGAGAGAGAGAGCGCGAG 4403
 2313 TGTGAG 2366
 4404 GCTGAAGCGCGAG 4457

RESULT 14
 US-09-249-585A-2
 ; Sequence 2, Application US/09249585A
 ; Patent No. 6417002
 ; GENERAL INFORMATION:
 ; APPLICANT: Horlick, Robert
 ; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
 ; FILE REFERENCE: 0867/0D905
 ; CURRENT APPLICATION NUMBER: US/09/249,585A
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 1926
 ; TYPE: DNA
 ; ORGANISM: Epstein Barr Virus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1926)
 ; OTHER INFORMATION: coding strand of EBNA-1 DNA
 US-09-249-585A-2

Query Match 2.3%; Score 119.2; DB 4; Length 1926;
 Best Local Similarity 53.4%; Pred. No. 5.2e-21;
 Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

1231 GTGATCAGAGCTGCTGCTGAGAGAGCGCTGCTCAGAGATGAGAGAGAGAGAGAGAGAG 1290
 515 GAG 574

1291 CTGCTGTGACATTTGAAGATTAAGAAAGCGGAGAACTTCAGCAGAGAGAGAGAGAGCTG 1350
 575 GGGCAG 634

1351 GAGAGCGCGCGCAAGCGCTCTTGAGCAGAGAGCGCAAAAGAGCAGAGAGCGGTTGCTCAG 1410
 635 GGGCAG 694

1411 CTGAGCGCGCGAG 1470
 695 CAG 754

1471 CAGCTGAGCTGAG 1530
 755 GGGCAG 814

1531 GAG 1590
 815 GAG 874

1591 CACTTGAAATGAG 1650
 875 CAG 934

1651 GAGGAG 1698
 935 GAGCAG 982

RESULT 15
 US-09-050-863-2
 ; Sequence 2, Application US/09050863
 ; Patent No. 611411
 ; GENERAL INFORMATION:
 ; APPLICANT: Lao, Ying

APPLICANT: Hwang, Betty
APPLICANT: Payan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 2.3%; Score 119.2; DB 3; Length 2580;
Best Local Similarity 53.4%; Pred No. 6,2e-21;
Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 1231 GTGATCAGAGGCTGCTGAGAGAGCCGTCGTCAGAGATGACGACGACGAGAAAGAA 1290
DB 898 GAGAGAGGAGCAGAGCAGAGAGAGGAGCAGAGAGGAGCAGAGAGAGAGAGAGAGAG 957
QY 1291 CTGCTGTGACATTTAAGTAAGAACCGGAGAACTTCAGAGAGGCACTGTGAGCTG 1350
DB 958 GGGCAGAGGGGAGCAGAGCAGAGAGAGGAGCAGAGCAGAGAGGAGCAGAGAGGAG 1017
QY 1351 GAGAAGGCGCCGCAAGGCTCTTGAGCAGCAGCAGAAAGCAGAGAGCGGTTGCTCAG 1410
DB 1018 GGGCAGAGAGCAGAGAGGAGCAGAGCAGAGAGAGGAGCAGAGAGGAGCAGAGAG 1077
QY 1411 CTGAGGCGCCGAGCAGAGAGAGAGAAAGCGGAGCGCCAGAGCAGAGAGCGCAAGCG 1470
DB 1078 CAGAGAGGAGCAGAGAGGAGCAGAGAGAGGAGCAGAGAGGAGCAGAGAGGAG 1137
QY 1471 CAGCTGAGCTGAGAGAGCAGCTGAGAGCAGAGAGAGGAGCTGAGAGCGCAGAGAGAG 1530
DB 1138 GGGCAGAGAGCAGAGAGGAGCAGAGAGGAGCAGAGAGGAGCAGAGAGGAGCAGAGAG 1197
QY 1531 GAGAGAGAGAGAGATCGAGAGCGCGCAGAGCGCAAAACGGGAACCTGAAAAGCAGCA 1590
DB 1198 GAGAGAGGAGCAGAGAGGAGCAGAGCAGAGAGAGGAGCAGAGAGGAGCAGAGAGGAG 1257
QY 1591 CAACCTGAATGAGAACGAGACCGAGAGCAGAGAACTCTGATCAGAGAGAAACAGAGCAG 1650
DB 1258 CAGAGAGGAGCAGAGAGGAGCAGAGAGGAGCAGAGAGGAGCAGAGAGGAGCAGAGAG 1317
QY 1651 GAGAGCACCCTGTGCTGTAAGGCAAGGAGAGAACTCTGAGATTGAG 1698
DB 1318 GAGCAGAGAGAGAGGAGCAGAGAGGAGCAGAGAGGAGCAGAGAGGAGGAG 1365

Search completed: March 14, 2003, 11:14:56
Job time : 488.483 secs

2:	gd_hng:	*
3:	gd_in:	*
4:	gb_om:	*
5:	gb_ov:	*
6:	gb_pat:	*
7:	gb_ph:	*
8:	gb.pl:	*
9:	gb.pr:	*
10:	gd_ro:	*
11:	gd_sts:	*
12:	gb_sy:	*
13:	gb_un:	*
14:	gb_vt:	*
15:	em_ba:	*
16:	em_fun:	*
17:	em_hum:	*
18:	em_in:	*
19:	em_nm:	*
20:	em_om:	*
21:	em_or:	*
22:	em_ov:	*
23:	em_pat:	*
24:	em_pl:	*
25:	em.pl:	*
26:	em.ro:	*
27:	em_sts:	*
28:	em_un:	*
29:	em_vt:	*
30:	em_hg_hum:	*
31:	em_hg_inv:	*
32:	em_hg_other:	*
33:	em_hg_mus:	*
34:	em_hg_pln:	*
35:	em_hg_rod:	*
36:	em_hg_mam:	*
37:	em_hg_vt:	*
38:	em_sy:	*
39:	em_hgo_hum:	*
40:	em_hgo_mus:	*
41:	em_hgo_other:	*

C	1	3721.4	73.2	3723	10	AF132478	AF132478 Mus muscu
	2	3639.4	71.6	5145	10	AF132481	AF132481 Mus muscu
	3	3529.4	69.4	5381	9	AF114488	AF114488 Homo sapi
	4	3517.2	69.2	4025	10	AF127798	AF127798 Rattus no
	5	3440.2	67.7	5287	9	AF064243	AF064243 Homo sapi
	6	3103.6	61.0	3812	10	AF133672	AF133672 Rattus no
	7	2875	58.5	6439	9	AF114487	AF114487 Homo sapi
	8	2896.4	57.0	7247	9	AF064244	AF064244 Homo sapi
	9	2047.2	40.3	4103	5	AF032118	AF032118 Xenopus l
	10	1404.2	27.6	2131	9	AK027846	AK027846 Homo sapi
	11	1255.6	24.6	3241	9	HS061166	U61166 Human SH3 d
	12	1110	21.8	1133	10	AF169621	AF169621 Mus muscu
C	13	1039.4	20.4	230937	2	AC126053	AC126053 Mus muscu
	14	883	17.4	1996	9	AF180522	AF180522 Homo sapi
C	15	813.6	16.0	78190	9	AC106760	AC106760 Homo sapi
	16	813.6	16.0	179264	2	AC023112	AC023112 Homo sapi
	17	731.4	14.4	5938	9	AB033082	AB033082 Homo sapi
	18	726.4	14.3	5828	9	AF248540	AF248540 Homo sapi
	19	724.2	14.2	4557	9	AF182139	AF182139 Homo sapi
	20	724.2	14.2	6091	9	AF182138	AF182138 Homo sapi
	21	693.4	13.6	136612	2	AC115529	AC115529 Rattus no
	22	692.6	13.6	180637	2	AC098552	AC098552 Rattus no
	23	677	13.3	4977	10	AF132479	AF132479 Mus muscu
	24	673.8	13.3	3594	10	AF132480	AF132480 Mus muscu
	25	659	13.0	4447	6	AX428899	AX428899 Sequence
	26	633.4	12.5	181782	2	AC116970	AC116970 Rattus no
	27	556.2	10.9	11777	9	AP000312	AP000312 Homo sapi
	28	556.2	10.9	100000	9	AP000193	AP000193 Homo sapi
	29	556.2	10.9	114928	9	AP000050	AP000050 Homo sapi
	30	556.2	10.9	151516	9	AP000117	AP000117 Homo sapi
	31	556.2	10.9	340000	9	AP000178	AP000178 Homo sapi
	32	553.8	10.9	2020	9	HS189075	AF064245 Homo sapi
	33	510	10.0	139113	2	AC129075	AC129075 Felis cat
C	34	509.4	10.0	235426	2	AC121074	AC121074 Canis fam
	35	508.6	10.0	178801	2	AC121075	AC121075 Canis fam
	36	484.8	9.5	2017	6	AX366305	AX366305 Sequence
	37	387.4	7.6	1329	9	AK021545	AK021545 Homo sapi
	38	375.2	7.4	529	9	HUM0P08B09	AF085873 Homo sapi
	39	333.2	6.6	2873	6	AR175271	AR175271 Sequence
	40	331.4	6.5	1391	9	AF001630	AF001630 Homo sapi
	41	331.4	6.5	2004	9	AF038189	AF038189 Homo sapi
	42	327.8	6.4	17160	9	AC012629	AC012629 Homo sapi
	43	258.8	5.1	4053	9	HS061167	U61167 Human SH3 d
	44	258	5.1	747	6	AR175268	AR175268 Sequence
C	45	230.2	4.5	175137	2	AC115339	AC115339 Rattus no

ALIGNMENTS

RESULT 1	AF132478				
LOCUS	AF132478	3723 bp	mRNA	linear	ROD 09-MAR-1999
DEFINITION	Mus musculus Esei protein mRNA, complete cds.				
ACCESSION	AF132478				
VERSION	AF132478.1	GI:4378884			
KEYWORDS					
SOURCE	Mus musculus.				
ORGANISM	Mus musculus.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus. 1 (bases 1 to 3723)				
AUTHORS	Seegar,A.S., Wang,W., Bishay,J., Cohen,S. and Egan,S.E.				
TITLE	The EH and SH3 domain Esei proteins regulate endocytosis by linking to dynamin and Eps15				

JOURNAL EMBO J. 18 (5), 1159-1171 (1999)
 MEDLINE 99164083
 PUBMED 10064583
 REFERENCE 2 (bases 1 to 3723)
 AUTHORS Sengar, A.S., Wang, W., Cohen, S., Bishay, J. and Egan, S.E.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAR-1999) Programs in Cancer & Blood
 Research/Developmental Biology, The Hospital for Sick Children, 555
 University Avenue, Toronto, ON M5G-1X8, Canada
 Location/Qualifiers

FEATURES
 source 1..3723
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 1..3642

CDS

/codon_start=1
 /product="Esel protein"
 /protein_id="A019746.1"
 /db_xref="GI:4378885"

/translation="MAQPTPTFGSLDVWVAIVVERAKHDQPSLKLPIAGITDDA
 RNFPSGLPQVLAQIMADNNNDGRNDQVFSIMKLIKQGLQPSLPPVM
 KQGPVASSAPAGIGIASMPETVAAPVWGSIPVMSPLVSAVPAVPLAN
 GAPVIAQPLPAFAHPAATWPKSSFSRSGPQLTKQKSPVSAVPAEVAVP
 OSSRLKYQLFNHDKTMSGLTGPARTILMOSSLPOKLASIMNLSIDDDGLTA
 EEFILAMHLIDVMSGQPLPVLPEYIPSPRRVSGSGMVISSSVDQLEPES
 SEDQOEKKLPVTFEDKRENNRSGVLEFRQALAEQCKEEDERLAQLERAEQER
 KEEREOQAKQLELEKOLEKORELEROREBERKETIRRAKRELERQLEBER
 NRQELNQRNKEQSGTVLKARKTLEFELALNDKHLQLEGLQDIRCLATROE
 IESTNSRELRIAEITHLOQLOESQOMGLRIPKQILSDQLKQONSLRDLIT
 LKALBAKELARQLEOLEDEVERETRSKLOEDIVNOMLEKREIHSKQOLKORSL
 EARLKOKEOERKSELLEKODEDAGRVOERDQWLEHVQOEORPRKPHEDLTKR
 EDSVRKEAEERAKPEMODKSLFPHQEPKALTOAAMSTTEKGLTISAQESVKY
 VYRALYEPRESHDEITIORGDIVWDSQGEQMLGELKGTGMPRANVAKIP
 ENENVTPAKPVDTLSAPAKLALRETPALPVTSSBETLNNMADSSTPSSNE
 KEBRDWDVMAQPSLTVPSAQLRORSAFTPAKTVGSSPSVLQGLQKRGVGLQAL
 YPWAKNDHNLNFKSDVITVEQDMWMAFQVOQKFPSPVYLKISGPVKSTSI
 DTPETSPASLKRVAAPAKPAIPGEFTAMTYTSSSEGDLTFQCGVIVTKKDG
 WMTGTGDSGVPSNYVRKIDSEGSGTGKSGKKEIADVIAAATPCBDLT
 APGOLILRRKNPGWEGELQARGKRIQGFNPANVYKLSPGNSKITPHELTAV
 OPACVIGIMMYDTANONDELAFSKQIILNVLNKEDPMKKEVSGVGLPFSNVKL
 TTDMPDSQQ"

BASE COUNT 1065 a 940 c 1026 g 692 t
 ORIGIN

Query Match 73.2%; Score 3721.4; DB 10; Length 3723;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 3722; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTTGAGATCTGGSCCAATCACTGTGAG 318
 DB 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTTGAGATCTGGSCCAATCACTGTGAG 60
 QY 319 GAAAGGCGCAAGCATGACGAGCTTCTAGGCTGAAGCGGATACGGGATTTATTA 378
 DB 61 GAAAGGCGCAAGCATGACGAGCTTCTAGGCTGAAGCGGATTTATTA 120
 QY 379 GGTATCAAGCGAGAACTTTTTCATCTGGGTTACTCAGCGCTGCTTAGACAA 438
 DB 121 GGTATCAAGCGAGAACTTTTTCATCTGGGTTACTCAGCGCTGCTTAGACAA 180
 QY 439 ATATGGCGCTTAGCGGACATGAATTAACGATGAGAGATGATCAAGTGAATTTTCCATA 498
 DB 181 ATATGGCGCTTAGCGGACATGAATTAACGATGAGAGATGATCAAGTGAATTTTCCATA 240
 QY 499 GCCATGAAGCTTATCAAACTGAAGCTAACAAGATATCAGTCCCTCCCACTTCCCTCC 558
 DB 241 GCCATGAAGCTTATCAAACTGAAGCTAACAAGATATCAGTCCCTCCCACTTCCCTCC 300
 QY 559 GTTCATGAACAGCAACCAAGTGTATTTTCCAGTGAACCAAGATTTGATAGAGGAGATT 618
 DB 301 GTTCATGAACAGCAACCAAGTGTATTTTCCAGTGAACCAAGATTTGATAGAGGAGATT 360
 QY 619 GCTAGATGCGACCACTGACAGCTGTTGCTCTGTCGCAATGGGCTCCATTCCAGTTGTT 678

DB 361 GCTAGCATGCCACCACTGACAGCTGTGCTCTCTGTCGCAATGGGCTCCATTCCAGTTGTT 420
 QY 679 GGAATGTCCTCCACCTTGTATGATCTTCTGCTCCCTCAGAGAGAGAGCTCCCTGGGCTAAC 738
 DB 421 GGAATGTCCTCCACCTTGTATGATCTTCTGCTCCCTCAGAGAGAGAGCTCCCTGGGCTAAC 480
 QY 739 GGGGCTCTCCCTGATACAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 798
 DB 481 GGGGCTCTCCCTGATACAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 799 AAGAGTCTCTCTGATGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 858
 DB 541 AAGAGTCTCTCTGATGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 QY 859 GCACAAATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 918
 DB 601 GCACAAATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 QY 919 TCAAGGCTGAAATACAGGAGTTATTCACAGCCACGACAAACTATGATGACACTT 978
 DB 661 TCAAGGCTGAAATACAGGAGTTATTCACAGCCACGACAAACTATGATGACACTT 720
 QY 979 ACAGGTCCTCCAGGAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 1038
 DB 721 ACAGGTCCTCCAGGAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
 QY 1039 TCAATATGAAATCTTTTGTGACATTCATGCAATGAAATCTCACTGACAGAAATTTATC 1098
 DB 781 TCAATATGAAATCTTTTGTGACATTCATGCAATGAAATCTCACTGACAGAAATTTATC 840
 QY 1099 CTAGTATGACACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158
 DB 841 CTAGTATGACACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 1159 CCAGAAATCATCT 1218
 DB 901 CCAGAAATCATCT 960
 QY 1219 AGCTTCTTCTTGTGTGATCAGAGCTGCTGAGAGCGCTGTCAGAGATGAGCAGCAG 1278
 DB 961 AGCTTCTTCTTGTGTGATCAGAGCTGCTGAGAGCGCTGTCAGAGATGAGCAGCAG 1020
 QY 1279 CCAGAGAAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
 DB 1021 CCAGAGAAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 QY 1339 AGTGTGAGCTGAGAGAGCGCGCCGCAAGCGCTCTTGGAGCAGCAGCAGCAGCAGCAG 1398
 DB 1081 AGTGTGAGCTGAGAGAGCGCGCCGCAAGCGCTCTTGGAGCAGCAGCAGCAGCAGCAG 1140
 QY 1399 CGGTTGGCTCAGCTGAGCGCGCGGACAGAGAGAAAGCGGAGCGCGCAGGAGCAG 1458
 DB 1141 CGGTTGGCTCAGCTGAGCGCGCGGACAGAGAGAAAGCGGAGCGCGCAGGAGCAG 1200
 QY 1459 GAGGCAAGCGGCAAGCTGAGCTGAGAGACAGCTGAGAGAGCGGAGCGGAGCTGAGCGG 1518
 DB 1201 GAGGCAAGCGGCAAGCTGAGCTGAGAGACAGCTGAGAGAGCGGAGCGGAGCTGAGCGG 1260
 QY 1519 CAGGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGCAGCGCCGCAAAACGGGAACTG 1578
 DB 1261 CAGGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGCAGCGCCGCAAAACGGGAACTG 1320
 QY 1579 GAAAGGAG 1638
 DB 1321 GAAAGGAG 1380
 QY 1639 AACAG 1698
 DB 1381 AACAG 1440
 QY 1699 TTAGAAGCTCTGATATGACAAAGAGCATCAGCTAGAGAGAAACTTTCAGAGATATCAGGTGT 1758
 DB 1441 TTAGAAGCTCTGATATGACAAAGAGCATCAGCTAGAGAGAAACTTTCAGAGATATCAGGTGT 1500

QY 1759 CGACTGGCAACCCAGAGCAAGAAATTGAGAGCAGCAAGCTTAGAGAGCTTAAGATT 1818
| | | | |
Db 1501 CGACTGGCAACCCAGAGCAAGAAATTGAGAGCAGCAAGCTTAGAGAGCTTAAGATT 1560
| | | | |
QY 1819 GCTGAATATACCCACTTACAGCAGCAGTTGACAGAAATCTCAGCAAAATGCTTGAAGACTT 1878
| | | | |
Db 1561 GCTGAATATACCCACTTACAGCAGCAGTTGACAGAAATCTCAGCAAAATGCTTGAAGACTT 1620
| | | | |
QY 1879 ATTCCAGAGAAACAGATACCTACAGTACCAAGTTAAACCAAGTCCAGAGCAAGCTTGGCAT 1938
| | | | |
Db 1621 ATTCCAGAGAAACAGATACCTACAGTACCAAGTTAAACCAAGTCCAGAGCAAGCTTGGCAT 1680
| | | | |
QY 1939 AGAGACTCGCTTCTTACCCCTCAAAAAGCCTTGAAGCAAAAGAGCTGGCCCGGACAG 1998
| | | | |
Db 1681 AGAGACTCGCTTCTTACCCCTCAAAAAGCCTTGAAGCAAAAGAGCTGGCCCGGACAG 1740
| | | | |
QY 1999 CTCGAGAGCAGCTGAGCAGAGTTGAGAGAGAGACCAAGTCAAAAGCTGACAGAGATTGAT 2058
| | | | |
Db 1741 CTCGAGAGCAGCTGAGCAGAGTTGAGAGAGAGACCAAGTCAAAAGCTGACAGAGATTGAT 1800
| | | | |
QY 2059 GTTTTCAACAACCAAGCTGAAGAACTGAGAGATACATAGCAAAAGCAACTCCAGAAAG 2118
| | | | |
Db 1801 GTTTTCAACAACCAAGCTGAAGAACTGAGAGATACATAGCAAAAGCAACTCCAGAAAG 1860
| | | | |
QY 2119 CAGAGGTCCTTGGAGGCGAGCACTGAAGCAGAAAAGAGCAGAGAGAGAAAGCCTGGAG 2178
| | | | |
Db 1861 CAGAGGTCCTTGGAGGCGAGCACTGAAGCAGAAAAGAGCAGAGAGAGAAAGCCTGGAG 1920
| | | | |
QY 2179 TTAGAGAGCAAAAAGAGAGACGCTCAGAGACGAGTTCAAGAAAAGGACAAAGATGCTTG 2238
| | | | |
Db 1921 TTAGAGAGCAAAAAGAGAGACGCTCAGAGACGAGTTCAAGAAAAGGACAAAGATGCTTG 1980
| | | | |
QY 2239 GAGCATGTGACAGCAGAGAGAGCAGCCAGCCCGGAAACCCACAGAGAGAGCAGACCTG 2298
| | | | |
Db 1981 GAGCATGTGACAGCAGAGAGAGCAGCCAGCCCGGAAACCCACAGAGAGAGCAGACCTG 2040
| | | | |
QY 2299 AAGAGGAAACAGAGTCAAGAAAGAGAGAGCGGAGAGAGAGAGAGCCGGAATGCA 2358
| | | | |
Db 2041 AAGAGGAAACAGAGTCAAGAAAGAGAGAGCGGAGAGAGAGAGAGCCGGAATGCA 2100
| | | | |
QY 2359 GACAAGCAGAGTGGCTTTTTCATCGCATAGAGCAGCTAAAGTGGCCACCCAGGCA 2418
| | | | |
Db 2101 GACAAGCAGAGTGGCTTTTTCATCGCATAGAGCAGCTAAAGTGGCCACCCAGGCA 2160
| | | | |
QY 2419 CCCCTGCTTACCAAGAGAAAGCCCGCTTACCATTTCTGACAGAGAGATGTAAGAGTG 2478
| | | | |
Db 2161 CCCCTGCTTACCAAGAGAAAGCCCGCTTACCATTTCTGACAGAGAGATGTAAGAGTG 2220
| | | | |
QY 2479 GTATATTACGAGGCTGTACCCCTTTGAATCCAGAAATCAGATGAGATCACATCCAG 2538
| | | | |
Db 2221 GTATATTACGAGGCTGTACCCCTTTGAATCCAGAAATCAGATGAGATCACATCCAG 2280
| | | | |
QY 2539 CCAGAGATATATAGTCATGTGTGATGAAGCCAGACTGAGAGCCAGAGATGGCTTGGAG 2598
| | | | |
Db 2281 CCAGAGATATATAGTCATGTGTGATGAAGCCAGACTGAGAGCCAGAGATGGCTTGGAG 2340
| | | | |
QY 2599 GAGCTGAAAGGAGAGAGAGAGTGGTCCCTGCAAACTATGAGAAAGATTTCCAGAAAT 2658
| | | | |
Db 2341 GAGCTGAAAGGAGAGAGAGAGTGGTCCCTGCAAACTATGAGAAAGATTTCCAGAAAT 2400
| | | | |
QY 2659 GAGGTTCCCACTCCAGCCCAACCAAGTACGATCTGACATCTGCCCCCTGCCCCCAACTG 2718
| | | | |
Db 2401 GAGGTTCCCACTCCAGCCCAACCAAGTACGATCTGACATCTGCCCCCTGCCCCCAACTG 2460
| | | | |
QY 2719 GCTCTGCTGAGAGCCCTGCTCTTTGCAAGTACCTTTCTGAGCCCTTCAACACCC 2778
| | | | |
Db 2461 GCTCTGCTGAGAGCCCTGCTCTTTGCAAGTACCTTTCTGAGCCCTTCAACACCC 2520
| | | | |
QY 2779 AACCACTGGGAGAGCTTGAAGTTCACAGTGGCCAGAGAGCTCAAAAGAGAGCCGAAAGC 2838
| | | | |
Db 2521 AACCACTGGGAGAGCTTGAAGTTCACAGTGGCCAGAGAGCTCAAAAGAGAGCCGAAAGC 2580
| | | | |
QY 2839 GACAACCTGGATACGCTGGGCGCTCAGCCTTCTCTGACCGTACTAGTGTGGCCAGTTA 2898
| | | | |
Db 2581 GACAACCTGGATACGCTGGGCGCTCAGCCTTCTCTGACCGTACTAGTGTGGCCAGTTA 2640
| | | | |
QY 2899 CGGAGAGATCAGCCTTTTACCCAGCAGCAGCCACTGGCTCTCCCATCTCCGCTCTG 2958
| | | | |
Db 2641 CGGAGAGATCAGCCTTTTACCCAGCAGCAGCCACTGGCTCTCCCATCTCCGCTCTG 2700
| | | | |
QY 2959 GGCAGAGGTGAAAAGGTGAAAGGCTCAAGCCGCAAGCCCTGTATCCCTGAGAGCCAAA 3018
| | | | |
Db 2701 GGCAGAGGTGAAAAGGTGAAAGGCTCAAGCCGCAAGCCCTGTATCCCTGAGAGCCAAA 2760
| | | | |
QY 3019 AAAGCAACCACTTAAATTTTAAACAAAAGTACGTCATCACCCTTCTGAAACGCAAGAC 3078
| | | | |
Db 2761 AAAGCAACCACTTAAATTTTAAACAAAAGTACGTCATCACCCTTCTGAAACGCAAGAC 2820
| | | | |
QY 3079 ATGTGCTGTTTGGAGAGTTCAAGGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAAA 3138
| | | | |
Db 2821 ATGTGCTGTTTGGAGAGTTCAAGGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAAA 2880
| | | | |
QY 3139 CTCATTTCAAGGCCCCGTAAAGAAATCCACAAAGCATGATCTGACCCTTAAAGTCTT 3198
| | | | |
Db 2881 CTCATTTCAAGGCCCCGTAAAGAAATCCACAAAGCATGATCTGACCCTTAAAGTCTT 2940
| | | | |
QY 3199 GCTAGTCTTAAAGAGATGCTTCCCGGCGCCCAAGCAGCATTTCCCGAGAAAGTCTT 3258
| | | | |
Db 2941 GCTAGTCTTAAAGAGATGCTTCCCGGCGCCCAAGCAGCATTTCCCGAGAAAGTCTT 3000
| | | | |
QY 3259 ATTGCAATGACATACAGAGATTTCTGAGCAAGAGATTTAACTTTCAAGCAAGGGAT 3318
| | | | |
Db 3001 ATTGCAATGACATACAGAGATTTCTGAGCAAGAGATTTAACTTTCAAGCAAGGGAT 3060
| | | | |
QY 3319 GTGATTGTGTTTACCAAGAAAGATGTGACTGTGAGCGGAAACGTTGGGAGCAAGTCT 3378
| | | | |
Db 3061 GTGATTGTGTTTACCAAGAAAGATGTGACTGTGAGCGGAAACGTTGGGAGCAAGTCT 3120
| | | | |
QY 3379 GAGATCTTCCCTTCTTAACTATGTAGAGCTTAAAGATTTCAAGAGGCTCTGGAACCTGCTG 3438
| | | | |
Db 3121 GAGATCTTCCCTTCTTAACTATGTAGAGCTTAAAGATTTCAAGAGGCTCTGGAACCTGCTG 3180
| | | | |
QY 3439 AAAACAGGAGTTTGAAGAAAAAACTGAAATTTGCGCAGTTATGCTTCTAGCTGCT 3498
| | | | |
Db 3181 AAAACAGGAGTTTGAAGAAAAAACTGAAATTTGCGCAGTTATGCTTCTAGCTGCT 3240
| | | | |
QY 3499 ACTGTGCCGAACCACTCAGCCCTGCTCTGAGCAGTGAATTTGATCCGAAAAAGAAC 3558
| | | | |
Db 3241 ACTGTGCCGAACCACTCAGCCCTGCTCTGAGCAGTGAATTTGATCCGAAAAAGAAC 3300
| | | | |
QY 3559 CCAAGTGTGATGTGGAGAGAGAACTGCAAGCTCGAGGAAAAAGGCCAATAGGCTG 3618
| | | | |
Db 3301 CCAAGTGTGATGTGGAGAGAGAACTGCAAGCTCGAGGAAAAAGGCCAATAGGCTG 3360
| | | | |
QY 3619 TTTCCAGCAAAATATGTCAAACTTTAAAGCCCGGAAACAGCAAAATCAGCCCAACTGAG 3678
| | | | |
Db 3361 TTTCCAGCAAAATATGTCAAACTTTAAAGCCCGGAAACAGCAAAATCAGCCCAACTGAG 3420
| | | | |
QY 3679 CTAACCAAGCCGAGTGTGAGCAGCAGCAGTGTGCAAGTATGAGGATGAGATTACAC 3738
| | | | |
Db 3421 CTAACCAAGCCGAGTGTGAGCAGCAGCAGTGTGCAAGTATGAGGATGAGATTACAC 3480
| | | | |
QY 3739 GCCCAGAACGATGAGCAACTAGCTTCAAGCAAAAGCCAGATATCAAGTCTTCAACAG 3798
| | | | |
Db 3481 GCCCAGAACGATGAGCAACTAGCTTCAAGCAAAAGCCAGATATCAAGTCTTCAACAG 3540
| | | | |
QY 3799 GAGGACCGGAGCTGAGTGAAGAGAGAGTCAAGTGGCAAGTTGGGCTCTTCCATCCAT 3858
| | | | |
Db 3541 GAGGACCGGAGCTGAGTGAAGAGAGAGTCAAGTGGCAAGTTGGGCTCTTCCATCCAT 3600
| | | | |
QY 3859 TATGTAAAGCTGACCAAGACATGAGACCCAGCAGCAATGATATATGTTGCTCATCC 3918
| | | | |
Db 3601 TATGTAAAGCTGACCAAGACATGAGACCCAGCAGCAATGATATATGTTGCTCATCC 3660
| | | | |
QY 3919 CCCCTCAGGCTTTGAAAGTCTTCAAGAGACCCACTATCTCCATATCAGTCCCAAGGGA 3978
| | | | |

```

Db      3661  CCCCCTGAGGCTTGAAGTCTCTCAAGAGACCCACTATCCCATATCATGCGCCAGAGGGA 3720
Qy      3979  TGA 3981
Db      3721  TGA 3723

RESULT 2
AF132481      5145 bp  mRNA  linear  ROD 09-MAR-1999
LOCUS      AF132481
DEFINITION Mus musculus Eseil protein mRNA, complete cds.
ACCESSION AF132481
VERSION    AF132481.1  GI:4378890
KEYWORDS
SOURCE     Mus musculus.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS    Sengar,A.S., Wang,W., Bishay,J., Cohen,S. and Egan,S.E.
TITLE      The EH and SH3 domain Ese proteins regulate endocytosis by linking
            to dynamin and Eps15
JOURNAL    EMBO J. 18 (5), 1159-1171 (1999)
MEDLINE    99164083
PUBMED     10064583
REFERENCE  2 (bases 1 to 5145)
AUTHORS    Sengar,A.S., Wang,W., Cohen,S., Bishay,J. and Egan,S.E.
TITLE      Direct Submision
JOURNAL    Submitted (02-MAR-1999) Programs in Cancer & Blood
            Research/Developmental Biology, The Hospital for Sick Children, 555
            University Avenue, Toronto, ON M5G-1X8, Canada

FEATURES
Source     1..5145
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            1..5145
            /codon_start=1
            /product="Eseil protein"
            /protein_id="AA019749.1"
            /db_xref="GI:4378891"
            /translation="MAQPTPPGSGSLDVAITVEERAHNDQFLSKPIAGFTGDQA
            RNFPGSGLPQVLAQIMALADNMNDMDQVFIAMK.I.LKIOGVLSTPPVM
            KOQVAFISSAPAFGIGIASMPITLAVAVPMGSIPIVGMSPILVSPVAPPLAV
            GAPVIOPLAPAPATAPWKSSEFSRSPGSLNTKOKASFPVAPAPAEVAPV
            OSSRLKYROLFNSHDKTMSGHLTGPQARTILMOSSLPOALASIMNISLIDDDKTLA
            EEPILAMHLIDVAMSGQPLPYLPPEYVPPSFRPRSSGMSVSSSVDDRLPERS
            SEDBOPEKKLVTFEDKRENFERSVELEKRAQLLEQKEDEERLAQLEKQER
            KEREROEAKROLEKOLEKERELEERERKELEERRAARLELEROLEWER
            NROELNORNEOEGTVLAKARKTLBEFLBALNDKGHLEGLKODIRCLAPORF
            IESTNSRELRIAEITHLQOOLQESQOMGLIPKQILSDQLKVOONSILHPSILT
            LKRALEAKELARQOLREOLDEVERETRSKLOLDIVENOLKEIRIHSKQOLORSL
            EAARLKOKEOERKSLLEKOKEDARQORROERDKMLLHVOCOEORPPEHEDIRK
            EDVSARKERERAKPEMDOKSRLLPHRDEPAKLTQAPMTTEGPIITISQSVKY
            VYPRALYPPRESHDEITLQPGDIVWDSOIGEGWAGGELKGFNPANVAKLP
            ENEVTPPAKPVTDLTSAAPKALRETPALPVTSSSESTIPNNMADSSWPSSNE
            KPEPTNMDTMAQPSLTVPSAGOLRORSATFPATGSGSPSVLQCGKVAOAL
            YEMRAKNDHNLNFKNSDVTIVLEQDDMMFQVCOQKMPSPSYKLSIGVRSKTSI
            DTGPSPASLKRVASPAKPAIPGEELFAMTYSSSGEGLTFQGGVIVVTKDD
            WMTGTVDGSGVFPNSVYRLKDSGSGTAMTSGSKGKPELADOUVASAAGPQDLT
            AGCOLILIRKNPGGMBEGLDARKKQKIQMFPRANYKLLSGTSKITPTEPRTLA
            QPACQVIGIMIDYIAONDELAFSGQIILVINKEDPMMWGEVSGQGLFSPSYKL
            TTDMDPSQMGCDLHLDMLPTERRKQGYIHLLVTEENYVNDLQVTELFQPLTE
            SLLTKEVAMIFVWKEKLIMCNILKILALRYKMSGEKMPVKMIGILSLOPHMG
            PYIRFSCOLNAAALIQOKTEADPDFKEFVKRLAMPDRCKGLSSFLTKMOVTRY
            PLIKNILENPNHSHSLKHALEKALECSOVNEGREKENSDELMTQAVOCC
            GLSBOLVENSVTNCLGPRKFLHSGLYAKSKKELEKGLPNDLILLITTPPSSGG
            DVFSPKSNLYQRYMTKPIFLNEVLVKLPDPSGDEPLFHLSHIDRVVTLAESINER
            TAWCVKTKAASELYIETEKREKAYLVARSQFATIGRLVNVVEGFEILKCRSHGS
            NYCEVTMSQCHITKTIQDTLNPKNNSQGFATIGLDEQEVLCITVEFLRDOFSDDEL
            GRTEIRVADIKKDGSKGVPVTKLHLHEVPTGEIIVVRDLQLFDEP"

BASE COUNT 1448 a 1298 c 1415 g 984 t
ORIGIN

```

```

Query Match      71.6%; Score 3639.4; DB 10; Length 5145;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      239  ATGGCTCAGTTTCCCAACACCTTTGGTGTGAGCTGATGTCTGGCCATPACTGTGAG 318
Db      1  ATGGCTCAGTTTCCCAACACCTTTGGTGTGAGCTGATGTCTGGCCATPACTGTGAG 60

Qy      319  GAAAGGCGCAAGCATGACACGAGCTTCTGCTGAAAGCCGATAGGGGATTTATTA 378
Db      61  GAAAGGCGCAAGCATGACACGAGCTTCTGCTGAAAGCCGATAGGGGATTTATTA 120

Qy      379  GGTGATCAAGCAGGAACTTTTTCATCTGGTAACTCCAGCTGTCTTACGAA 438
Db      121  GGTGATCAAGCAGGAACTTTTTCATCTGGTAACTCCAGCTGTCTTACGAA 180

Qy      439  ATATGGGGCTTAGGGGACATGAAATTAACGATGGAAGATGATCAAGTGAATTTCCATA 498
Db      181  ATATGGGGCTTAGGGGACATGAAATTAACGATGGAAGATGATCAAGTGAATTTCCATA 240

Qy      499  GCCATGAAGCTTATCAAACTGAACTCAAGATATCAGCTCCCTCCACACTTCCCT 558
Db      241  GCCATGAAGCTTATCAAACTGAACTCAAGATATCAGCTCCCTCCACACTTCCCT 300

Qy      559  GTCATGAAGCAGCAACCACTGCTATTTCCAGTGCACCAAGATTTGATAGAGGAT 618
Db      301  GTCATGAAGCAGCAACCACTGCTATTTCCAGTGCACCAAGATTTGATAGAGGAT 360

Qy      619  GCTGACATGACCACTCAACGCTGTTGCTCCGTGCAATGGGCTCATTCAGTGT 678
Db      361  GCTGACATGACCACTCAACGCTGTTGCTCCGTGCAATGGGCTCATTCAGTGT 420

Qy      679  GGAATGTCCTCAACCTTAGTATCTTCTGCTCCCTCCAGACAGAGTCCCTGGCTAAC 738
Db      421  GGAATGTCCTCAACCTTAGTATCTTCTGCTCCCTCCAGACAGAGTCCCTGGCTAAC 480

Qy      739  GGGGCTCTCCCGCATACAGCCCTCTGCTGCTGGGATCTGAGGACATAGGCCA 798
Db      481  GGGGCTCTCCCGCATACAGCCCTCTGCTGCTGGGATCTGAGGACATAGGCCA 540

Qy      799  AAGATGTCCTTCCACAGAGTGTGCTGCAAGGCTCAATTAACAATAAGTACAGAG 858
Db      541  AAGATGTCCTTCCACAGAGTGTGCTGCAAGGCTCAATTAACAATAAGTACAGAG 600

Qy      859  GCAATATCATTCATATGTCGCCAGGCGCCCTCCAGACAGAAATGGCTGTGCTCA 918
Db      601  GCAATATCATTCATATGTCGCCAGGCGCCCTCCAGACAGAAATGGCTGTGCTCA 660

Qy      919  TCAAGGCTGAAATACAGGAGTTTATCAAGCCAGCAAGCAAACTATGATGAGCACT 978
Db      661  TCAAGGCTGAAATACAGGAGTTTATCAAGCCAGCAAGCAAACTATGATGAGCACT 720

Qy      979  ACAGGTCCTCCAGGCAAGCACTATTCATGCAATCAAGTTTACCCAGGCTCAGTGG 1038
Db      721  ACAGGTCCTCCAGGCAAGCACTATTCATGCAATCAAGTTTACCCAGGCTCAGTGG 780

Qy      1039  TCAATATGGAATCTTTTGTGACATTGATCAAGATGGAATCTCACTGAGAGAAATTTAT 1098
Db      781  TCAATATGGAATCTTTTGTGACATTGATCAAGATGGAATCTCACTGAGAGAAATTTAT 840

Qy      1099  CTAGCTATGACCTAATATGATGTTGCCATGCTGATCAGCACTGCGCGCTCTGCT 1158
Db      841  CTAGCTATGACCTAATATGATGTTGCCATGCTGATCAGCACTGCGCGCTCTGCT 900

Qy      1159  CCAGAAATACATCTCTCTCTTCAAGAGATTCGCTCCGCGAGTGGATGTCCGTCATA 1218
Db      901  CCAGAAATACATCTCTCTCTTCAAGAGATTCGCTCCGCGAGTGGATGTCCGTCATA 960

Qy      1219  AGCTCTTCTCTGATGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGATGAGAGCGAG 1278
Db      961  AGCTCTTCTCTGATGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGATGAGAGCGAG 1020

```



```

Db 3181 AAAACAGGAGTTAGGAAAAAACTGAAATTGCCCCAGTTATTCCTCTCAAGCTGCT 3240
Qy 3499 ACTGTCCTCCGACAACTCACTCCCTGGCTCTCTGGGACCTGATTCGATCCGAAAAAAGAAC 3558
Db 3241 ACTGTCCTCCGACAACTCACTCCCTGGCTCTCTGGGACCTGATTCGATCCGAAAAAAGAAC 3300
Qy 3559 CCGAGGTGATGATGAGGAGGAGAACTGCAAGCTCGAGGGGAAAAAGCCGAGATAGGGTGG 3618
Db 3301 CCGAGGTGATGATGAGGAGGAGAACTGCAAGCTCGAGGGGAAAAAGCCGAGATAGGGTGG 3360
Qy 3619 TTTCAGCAATTATGTCAACTTCTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3678
Db 3361 TTTCAGCAAAATTATGTCAACTTCTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3420
Qy 3679 CTACCCCAAGCCGACGATGCGACGACGATGTCGAGTATCGGATGTAACGATTACACC 3738
Db 3421 CTACCCCAAGCCGACGATGCGACGACGATGTCGAGTATCGGATGTAACGATTACACC 3480
Qy 3739 GCCCAGACGATGACGAACTAGAGCTTCAGCAAGGCGACATCATCAAGCTCCCAACGAG 3798
Db 3481 GCCCAGACGATGACGAACTAGAGCTTCAGCAAGGCGACATCATCAAGCTCCCAACGAG 3540
Qy 3799 GAGGACCCGACGATGTCGAAAAAGAGATGAGGCGAAAGTTGGCTCTTCCCATCAAT 3858
Db 3541 GAGGACCCGACGATGTCGAAAAAGAGATGAGGCGAAAGTTGGCTCTTCCCATCAAT 3600
Qy 3859 TATGTAAGCTGACCAAGCATGAGCCCGACGACCAATG 3899
Db 3601 TATGTAAGCTGACCAAGCATGAGCCCGACGACCAATG 3641

RESULT 3
AF114488 5381 bp mRNA linear PRI 16-JUL-2002
LOCUS AF114488 Homo sapiens interseccion short isoform (ITSN) mRNA, complete cds.
DEFINITION AF114488
ACCESSION AF114488.1 GI:4808824
VERSION AF114488.1
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 5381)
AUTHORS Pucharcos, C., Fuentes, J. J., Cases, C., de la Luna, S., Alcantara, S.,
Arbones, M. L., Soriano, E., Estivill, X. and Pritchard, M.
Alu-splice cloning of human interseccion (ITSN), a putative
multivalent binding protein expressed in proliferating and
differentiating neurons and overexpressed in Down syndrome
cell lines.
JOURNAL Eur. J. Hum. Genet. 7 (6), 704-712 (1999)
MEDLINE 99415290
PUBMED 10482960
REFERENCE 2 (bases 1 to 5381)
AUTHORS Pucharcos, C., Fuentes, J. J., Pritchard, M. and Estivill, X.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1998) Medical and Molecular Genetics Center,
Cancer Research Institute, U/Hospital de Llo., Avia.
Castelldefels km. 2,7, Barcelona 08907, Spain
FEATURES
source
1..5381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
1..5381
/gene="ITSN"
/feature="ITSN"
269..3931
/gene="ITSN"
/codon_start=1
/product="interseccion short isoform"
/protein_id="AAD29953.1"
/db_xref="GI:4808825"
/translation="MAQFPTPGSLDIWATVEBRAGHDOFSLKISGFTGDOA
RMFFQSGLPQVLAQIWMALDMNNDGMDQVEFIAMKLIKLGQVLSALPVM"

```

```

KOOPVAISSAPARFGMGISMPBTRVAPVPMGSIPIVGMSPITVSSVPTAAPPILAN
GAPVLPPLPAPFAPATLIPKSSSRSSPSGSLTKLQAKQSPVAVSPAPVAVP
QSSRLKRYRLPNSHDKTMSGLTGPARTILMOSSLIPQOLASINLSDIDDGKTLTA
BERILAMHLDIVAMSGOLPVPVPEPIPPSRFRVSGSISVISTSDOLPEBPV
LEDEOOLBKULPVTPEDEKRENFERKLEERKQALIEQKQERLQALERAEQOE
KREBEROEERKRLERKLEERKLEERKLEERKLEERKLEERKLEERKLEERKLE
RNRROBLNORNMDEDIVLAKAKKTLFEELIELELDKHOGLERLODRGLTTORO
EIESTRKSLRELRIAETHTHLOOLESQOMLEALIEPKOILNOLKONONSLHRSIV
TLKRALERKLELARDOLDEVERKTRKLEIDILPNQLKELERHKOOLQOKS
MEARERKQERKLEERKLEERKLEERKLEERKLEERKLEERKLEERKLEERKLE
KREESVKKQDEBERKQERKLEERKLEERKLEERKLEERKLEERKLEERKLEERK
KVYVYALYEPESRSHDEITIOGDIIVYKGVESOTGEBGMGLKGTGWFPA
NVAEKIENEBEPAPVPTVSTAPAPKALBETPAPLAVTSESTPEPNNMADSSST
MPTSTKEPTEOTNMWMAAPSLTVPSAGOLBORSAPFPAATATGSSPSVYGOQKE
GLOAQOLIPWRKNDHNLNFKNDIVTVIEQDDMFRGVQOKHPSPVYKILSGP
IRKSTMSDSSSPASLRVASPAKPVASBEPFLAMTVSSSQDGLTFQGGVIL
VTBKDDMDWTGVGVKAVPSPNYVRLKDSBSGAGKGSJGKRELAQVIASTAT
GPEQLTLAPGQILIRKKNPGMWBEGLDARKKROIGWFPANVYKLSPGSKITPT
EPKSTALAAACOVIMDYTAQNDDDELAFNKGQILINVLNKEDPMMWKGEVNGVGLF
PSNYVRLTITMDPSQO"
BASE COUNT 1629 a 1192 c 1304 g 1256 t
ORIGIN
Query Match 69.4%; Score 3529.4; DB 9; Length 5381;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;
Qy 7 GAGAGAGAGTGAAGGCGCGGAGAGGCGCGGACGCTTGTCTCCCTAGTACGGCGCT 66
Db 27 GAGAGAGAGTGAAGGCGCGGAGAGGCGCGGACGCTTGTCTCCCTAGTACGGCGCT 86
Qy 67 CGCAAGAGAGATCCCGAGCGGCTCCCGAGCGCGGAGCGGAGCGAGCGAGCGCGGCG 126
Db 87 CGCAAGAGAGATCCCGAGCGGCTCCCGAGCGCGGAGCGGAGCGGAGCGGCGGCG 135
Qy 127 GGGATGCTGCGCGGCTGCGGACTGGCGCTTCTGCGC-CCGCGCTGCGGCTGCACTGA 185
Db 136 GGGATGCTGCGCGGCTGCGGACTGGCGCTTCTGCGC-CCGCGCTGCGGCTGCACTGA 195
Qy 186 TTTGTGAGAGGCGCGGCGCGGCGCGGCGGAGATGAGGCGTGCATGACGAGATGA 245
Db 196 TTTGTGAGAGGCGCGGCGCGGCGCGGCGGAGATGAGGCGTGCATGAGAGATGA 255
Qy 246 ACGTAATAGAACCATGCGCTCAGTTTCCACACCTTTGGTGTAGCTGTGAGCTGGGC 305
Db 256 AAGTAACAGAACCATGCGCTCAGTTTCCACACCTTTGGTGTAGCTGTGAGCTGGGC 315
Qy 306 CATACCTGTGAGAGAGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 365
Db 316 CATACCTGTGAGAGAGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 375
Qy 366 GGGATTTTATTAGTGTGATCAAGGAGGAACTTTTTCATCAATCTGGGTTTACCTGAGC 425
Db 376 TGGATTTTATTAGTGTGATCAAGGAGGAACTTTTTCATCAATCTGGGTTTACCTGAGC 435
Qy 426 TGTCTTGACCAAAATATGCGCGCTAGCGGACATGATTAAGATGAGATGATCAAGT 485
Db 436 TGTCTTGACCAAAATATGCGCGCTAGCGGACATGATTAAGATGAGATGATCAAGT 495
Qy 486 GGAATTTTCCATGAGGCGGAGGAGGCTTATCAAACTGAGGCTGCAAGATATGAGTCCCTC 545
Db 496 GGAATTTTCCATGAGGCGGAGGAGGCTTATCAAACTGAGGCTGCAAGATATGAGTCCCTC 555
Qy 546 CACACTTCCCTCTGTCATGAAAGAGCAACAGTGGCTATTTCCAGTGCACACAGATTTGG 605
Db 556 TGCACTTCCCTCTGTCATGAAAGAGCAACAGTGGCTATTTCCAGTGCACACAGATTTGG 615
Qy 606 TATAGGAGGATGCTGATGACCACTGACCACTGACCTGCTGCTGCTGCTGCTGCTGCTG 665
Db 616 TATAGGAGGATGCTGATGACCACTGACCACTGACCTGCTGCTGCTGCTGCTGCTGCTG 675
Qy 666 CATCCAGTGTGATGATGCTCAACCTTAGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 725

```

Dh 676 CATTCAGTTGTTGAATGTCCTCAACCTTAGTATCTTCTGCCACAGCAGCTGTGCC 735
Qy 726 TCCCTGGTAAACGGGGGCTCTCCGCTCATPACGCTCGCTGGGTTGGCATCTGCC 785
Db 736 CCCCCTGGCTAAAGGGGCTCCCTCTGTATACAACTCTGCTGCACTTGTCTCATCTGCC 795
Qy 786 AGCCATGAGGCAAGAGTCTTCTTCCAGAGATGCTGCTCCAGAGGTCACAATTAACAC 845
Db 796 AGCCATGAGGCAAGAGTCTTCTTCCAGAGATGCTGCTCCAGAGGTCACAATTAACAC 855
Qy 846 TAAATTACAAAGGCAAGTCAATTTGATGTGGCAAGTGTCCCAAGTGGCAGAGTGGCC 915
Db 856 TAAATTACAAAGGCAAGTCAATTTGATGTGGCAAGTGTCCCAAGTGGCAGAGTGGCC 915
Qy 906 TGTGCTCTGCTATCAAGGCTGAATATACAGGAGTATTCACAGCAGCAAGAACTAT 965
Db 916 TGTGCTCTGCTATCAAGGCTGAATATACAGGAGTATTCACAGCAGCAAGAACTAT 975
Qy 966 GAGTGACACTTAAACAGGCTCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCCA 1025
Db 976 GAGTGACACTTAAACAGGCTCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCCA 1035
Qy 1026 GAGTGACACTTAAACAGGCTCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCCA 1085
Db 1036 GAGTGACACTTAAACAGGCTCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCCA 1095
Qy 1086 AGAAGAAATTTATCTTGAATGCAATCTTCTGATGCAATGATGAAAGTCACTGCC 1145
Db 1096 AGAAGAAATTTATCTTGAATGCAATCTTCTGATGCAATGATGAAAGTCACTGCC 1155
Qy 1146 GCCCGTCTGCTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGCGAGTGG 1205
Db 1156 ACCCTGCTGCTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGCGAGTGG 1215
Qy 1206 GATGTCGCTGATGAGTCTTCTTCTGATGCAAGGCTGCTGAGAGCGCTGCTGAGA 1265
Db 1216 TATATGCTGATGAGTCTTCTTCTGATGCAAGGCTGCTGAGAGCGCTGCTGAGA 1275
Qy 1266 GATGTCGCTGATGAGTCTTCTTCTGATGCAAGGCTGCTGAGAGCGCTGCTGAGA 1322
Db 1276 GATGTCGCTGATGAGTCTTCTTCTGATGCAAGGCTGCTGAGAGCGCTGCTGAGA 1335
Qy 1323 GAACTTGAGCAGGAGGCTGAGAGGCTGAGAGCGCGCGCAAGCGCTTGGAGAGCA 1382
Db 1336 GAACTTGAGCAGGAGGCTGAGAGGCTGAGAGCGCGCGCAAGCGCTTGGAGAGCA 1395
Qy 1383 GCCCAAAAGCAGGAGCGGCTGCTGAGAGCGCGCGCAAGCGCTTGGAGAGCA 1442
Db 1396 GCCCAAAAGCAGGAGCGGCTGCTGAGAGCGCGCGCAAGCGCTTGGAGAGCA 1455
Qy 1443 GGAAGCGCAGAGAGCGGCTGCTGAGAGCGCGCGCAAGCGCTTGGAGAGCA 1502
Db 1456 GGAAGCGCAGAGAGCGGCTGCTGAGAGCGCGCGCAAGCGCTTGGAGAGCA 1515
Qy 1503 GCGGAGCTGAGAGCGGCTGAGAGGAGAGGAGAGGAGAGTGCAGAGGCGCGAGCG 1562
Db 1516 GCGGAGCTGAGAGCGGCTGAGAGGAGAGGAGAGGAGAGTGCAGAGGCGCGAGCG 1575
Qy 1563 CGCAAAACGGGAACTGGAAGGCGAGCAACTTGAATGGGAAACGAAACCGGAGACAGA 1622
Db 1576 TGCAAAACGGGAACTTGAAGGCGAGCAACTTGAATGGGAAACGAAACCGGAGACAGA 1635
Qy 1623 ACTCTGAATCAGAGAAACAGAGAGGAGGAGCGCTGCTGAGAGGAGAGAGGAA 1682
Db 1636 ACTCTGAATCAGAGAAACAGAGAGGAGGAGCGCTGCTGAGAGGAGAGAGGAA 1695
Qy 1683 GACTCTGAGTGTGATGAGAGCTGAGAGCAAAAGAGTCAAGTATGAGAGGAACT 1742
Db 1696 GACTCTGAGTGTGATGAGAGCTTGAATGATGAGAGGAGTAACTGAGAGGAACT 1755
Qy 1743 TCAGAGATATCAGTGTGAGTGCAGCCAGAGGCAAGAAATGGAGAGCAAGCAAGTC 1802
Db 1756 TCAGAGATATCAGTGTGAGTGCAGCCAGAGGCAAGAAATGGAGAGCAAGCAAGTC 1815

Qy 1803 TAGAGACTAAGAAATTCGTGAATCAACCACTTAACAGCAGCAGTTCGAGAACTCAGCA 1862
Db 1816 TAGAGAGTTGAATTTCCGAAATCACCCATCTACAGCAACAAATTACAGAAATCTCAGCA 1875
Qy 1863 AATGCTTGAAGACTTTTTCAGAGAAACAGATATCTAGTACAGTAAACAAAGTCCA 1922
Db 1876 AATGCTTGAAGACTTTTTCAGAGAAACAGATATCTAGTACAGTAAACAAAGTCCA 1935
Qy 1923 GCAGAAACAGTTTGCATGAGAGTGTCTTCAACCTCAAAAGAGCTTGGAGCAAAAGCA 1982
Db 1936 GCAGAAACAGTTTGCATGAGAGTGTCTTCAACCTCAAAAGAGCTTGGAGCAAAAGCA 1995
Qy 1983 GCTGAGCCGCGAGCAGGCTCCGAGAGCAGTGCAGAGGAGAGAGAGCAAGTCAAA 2042
Db 1996 ACTAGCTGCGCAGCAGCTCAAGAGCAAACTGAGTGAAGTGAAGAAAGAACTAGATCAAA 2055
Qy 2043 GCTGAGAGAGTGTATGTTTTCACAAACAGCTGAGAGAACTGAGAGATATCTAGCA 2102
Db 2056 ACTAGAGAGATGATATTTTCAATTAATCAGCTGAGAGAACTAAGAGAAATACAAATTA 2115
Qy 2103 ACAGCACTCCAGAGAGAGAGGCTCCGAGAGCGAGCTGAAGGAGAGAGAGCAGCA 2162
Db 2116 GCAGCACTCCAGAGAGAGAGGCTCCGAGAGCGAGCTGAAGGAGAGAGAGCAGCA 2175
Qy 2163 GAGGAGAGCCTGAGAGTGAAGAGCAAAAGAGAGAGCTCAGAGAGAGTTCAGAAAG 2222
Db 2176 AGGAAAGATCATAGAAATTTAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAG 2235
Qy 2223 GAGCAAGATGCTGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2279
Db 2236 GAGCAAGATGCTGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2295
Qy 2280 CCAG 2339
Db 2296 CCAG 2355
Qy 2340 AGCCAGCCGAGAAATGCAAGCAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGCAGC 2399
Db 2356 AGCCAGCCGAGAAATGCAAGCAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGCAGC 2415
Qy 2400 TAAAGTGGCCACCCAGAGCAGCTGCTGATCAACAGAGAAAGCCGCTTACATTTCTGC 2459
Db 2416 TAAAGTGGCCACCCAGAGCAGCTGCTGATCAACAGAGAAAGCCGCTTACATTTCTGC 2475
Qy 2460 ACAGAGAGTGTAAAGTGTGTATTAACCGAGCGCTGATACCTTGAATCAGAGAGCA 2519
Db 2476 ACAGAGAGTGTAAAGTGTGTATTAACCGAGCGCTGATACCTTGAATCAGAGAGCA 2535
Qy 2520 CGATGAGATCAACATCCAGCAGAGAGATATAGTAT -----GCTGATGA 2564
Db 2536 TATATGAATCACTATCCAGCAGAGAGATATAGTATTAAGGAGAAATGGGTGATGA 2595
Qy 2565 AAGCCAGACTGAGAGAGCAGAGATGCTTGAAGAGAGCTGAAGAGAGAGAGAGAGTGT 2624
Db 2596 AAGCCAGACTGAGAGAGCAGAGATGCTTGAAGAGAGCTGAAGAGAGAGAGAGTGT 2635
Qy 2625 CCTTGAAGTATGAGAGAAAGATTTCCAGAAATGAGGTTTCCACTCCAGCAGAAACAGT 2684
Db 2656 CCTTGAAGTATGAGAGAAATTTCCAGAAATGAGGTTTCCACTCCAGCAGAAACAGT 2715
Qy 2685 GACCGATGAGATGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2744
Db 2716 GACCGATGAGATGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2775
Qy 2745 GCGAGTGAATCTTGTGAG 2804
Db 2776 GCGAGTGAATCTTGTGAG 2835
Qy 2805 GTGGCCAGAGAGCTCAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2864
Db 2836 GTGGCCAGAGAGCTCAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2895

QY 2865 GCGCTTCTGACCGTACCTAGTGTGCGCAGTATACGAGATCAGCCTTTACCCACG 2924
 Db 2896 GCGCTTCTGACCGTACCGTACCGCAGTATACGAGATCAGCCTTTACCCACG 2955
 QY 2925 CAGAGCCATCTGCTCTCCCACTCTCCGCTCTGCGCAGGCTGAAAGTGAAGGCT 2984
 Db 2956 CAGAGCCATCTGCTCTCCCACTCTCCGCTCTGCGCAGGCTGAAAGTGAAGGCT 3015
 QY 2985 ACAAGCGCAAGCCTGTATCCCTGAGAGCCAAAGAACCACTTAATTTTACAA 3044
 Db 3016 ACAAGCTCAAGCCTGTATCCCTGAGAGCCAAAGAACCACTTAATTTTACAA 3075
 QY 3045 AAGTGAAGTATACCGTCTTGAAGACAGACATGTGTGTTTGAAGATTTCAAG 3104
 Db 3076 AAGTGAAGTATACCGTCTTGAAGACAGACATGTGTGTTTGAAGATTTCAAG 3135
 QY 3105 TCAGAGGGTGTGTTCCCAAGTCTACGTGAACATCTTACGGCCCTGAAGAAATC 3164
 Db 3136 TCAGAGGGTGTGTTCCCAAGTCTACGTGAACATCTTACGGCCCTGAAGAAATC 3195
 QY 3165 CACAGCATCTGATCTGCGCTACTGAAAGTCTGCTAGCTTAAGAGAGTCTCCCC 3224
 Db 3196 TACAGCATCTGATCTGCGCTACTGAAAGTCTGCTAGCTTAAGAGAGTCTCCCC 3255
 QY 3225 GCGCCGCAAGCCAGCCATTCGCGGAGAGATTTATTCGATGTACATACGAGATTC 3284
 Db 3256 AGCAGCCAGCCAGCCATTCGCGGAGAGATTTATTCGATGTACATACGAGATTC 3315
 QY 3285 TGAGCAAGAGATTTTAACTTTCAGCAAGGGAGTGTATTTGTGTTACCAAGAAATG 3344
 Db 3316 TGAGCAAGAGATTTTAACTTTCAGCAAGGGAGTGTATTTGTGTTACCAAGAAATG 3375
 QY 3345 TGACTGTGACGAGAAAGTGTGCGCAGCAAGTCCGAGTCTTCCCTTACTATGTGAG 3404
 Db 3376 TGACTGTGACGAGAAAGTGTGCGCAGCAAGTCCGAGTCTTCCCTTACTATGTGAG 3435
 QY 3405 GCTTAAAGATTCAGAGGCTCTGAACTGTGAGAAACAGGAGATTTGAGAAAAACC 3464
 Db 3436 GCTTAAAGATTCAGAGGCTCTGAACTGTGAGAAACAGGAGATTTGAGAAAAACC 3495
 QY 3465 TGAATTTGCGCAGATTTATGCTTCTGAGGTGCTACTGCTCCGCAACATCAGCTGCG 3524
 Db 3496 TGAATTTGCGCAGATTTATGCTTCTGAGGTGCTACTGCTCCGCAACATCAGCTGCG 3555
 QY 3525 TCTGTGCGCAGCTGATTTCTGATCCGAAAAAGAACCCAGGTGATGTGTGAGAGAACT 3584
 Db 3556 CCTGTGTGCGCAGCTGATTTCTGATCCGAAAAAGAACCCAGGTGATGTGTGAGAGAACT 3615
 QY 3585 GCAAGCTCGAGGAAAGACGCGCAGATAGGCTGTTCCAGAAATTAATGTCAAACTTCT 3644
 Db 3616 GCAAGCTCGAGGAAAGACGCGCAGATAGGCTGTTCCAGAAATTAATGTCAAACTTCT 3675
 QY 3645 AAGCCCCGGAACAGCAAAATCACCCCACTGAGTACCAAGACGCGTCAAGCCAGC 3704
 Db 3676 AAGCCCCGGAACAGCAAAATCACCCCACTGAGTACCAAGACGCGTCAAGCCAGC 3735
 QY 3705 AGTGTGCGCAGGTGATGGGATGTACATTAACCGCCAGAACGATGAGAACTAGCCTT 3764
 Db 3736 AGTGTGCGCAGGTGATGGGATGTACATTAACCGCCAGAACGATGAGAACTAGCCTT 3795
 QY 3765 CAGCAAGGCGCAGATCATCAAGTCTCTCAACAAAGAGAGCCGCGACTGGTGAAGAGGA 3824
 Db 3796 CAGCAAGGCGCAGATCATCAAGTCTCTCAACAAAGAGAGCCGCGACTGGTGAAGAGGA 3855
 QY 3825 AGTCACTGAGCAAGTGTGGCTCTTCCCATCAATTAATGTAAAGTGAACAGACATGGA 3884
 Db 3856 AGTCACTGAGCAAGTGTGGCTCTTCCCATCAATTAATGTAAAGTGAACAGACATGGA 3915
 QY 3885 CCCCAGCCAGCAATGATATATGTGTCATCCTCCCTCAGGCTTGAAGTCTCTCAA 3944
 Db 3916 CCCCAGCCAGCAATGATATATGTGTCATCCTCCCTCAGGCTTGAAGTCTCTCAA 3975
 QY 3945 GAGACCACTATCCATCATCACTGCGCAGAGGATGATGGAGATGCAAGCTTGTATCATG 4004

Db 3976 GAGACCACTATCCCATATCATCTGCCAGAGGGATGATGGAGATGCAAGCTTGATCATG 4035
 QY 4005 TGACTTGCAGATATATCACTATGCTCTTCTGAGTAGAANAATCACTCATGAGAGCATTT 4064
 Db 4036 TGACTTGCAGATATATCACTATGCTCTTCTGAGTAGAANAATCACTCATGAGAGCATTT 4095
 QY 4065 ACCCTATTGACCTTATGATCATGTGATCGAAATGTCTGATGACGCGCAGAGAGCAG 4124
 Db 4096 ACCCTATTGACCTTATGATCATGTGATCGAAATGTCTGATGACGCGCAGAGAGCAG 4155
 QY 4125 AAGC - AATTGCAAGCTGCAAGAGGTGTGCTGCTTTGGGGCTTCTAGTCACTC 4182
 Db 4156 GAGCAAAATTTACAAAACACAGAGGTAGTGGCTCTTTTGGGGCTTCTAGTCACTC 4215
 QY 4183 AGACTGACCGG - CCGCGCTTCAACAGGCGGCTTCAATGTGTTTAAGATTAATTTTAA 4241
 Db 4216 AATTGACTTTCCCGCAGCTTGTGACAGGTGCTTCAATGTGTTTAAGATTAATTTTAA 4275
 QY 4242 TGTGTATTTTACCTTTTAAATTAATCTCAATCAATTAATTTTGTGCTATTTGGTTT 4301
 Db 4276 TATATATTTTACCTTTTAAATTAATTAATTAATTAATTTTGTGCTATTTGGTTT 4335
 QY 4302 TACAAAAACCCACTATCAAGAGTGTCTGTGCGACGATTAATAATGCTGCTCGG 4361
 Db 4336 TGCAAAAAGCCACTATCAAGAGATG - - - - CTGATGTGCTATTAATAATTTTCCAA 4390
 QY 4362 GATACCGTAACCTGAGAGCTGTGCTGATCTTTCGCTTTCAGTGTCCCAACACAT 4421
 Db 4391 ATGTCCATTAATCTGAGACTGTGATGTTTTCATTTTTCAGTGTCCCAACAT 4450
 QY 4422 TGTGTATTTTGGGCTGTTCC - - - - TGCGTAGACACAGAGAGATGGGTACTCTGT 4477
 Db 4451 TGTGCACTTTGGGCTTTCCTTCCCTTAACATAGAGAGCAGAGAGTCAATCTCTGT 4510
 QY 4478 TTTTAAATGTGTATGTAGACTGACCTGACTATGAGAGGGCTTATGCTGTGTGAC 4537
 Db 4511 TTTTAAAGCGTATTA - GAATGAGCCCAATTAAGCGAAGGTGTTGTGTTGTGTGT 4569
 QY 4538 ATCAAGTGTAC - TGTGCGCATGTATCAATCTGTACCGAAGAGTGTGT - CTTCCTCCAT 4595
 Db 4570 ATCAAGTGTACCTTGTGTGAGATGTATATCTGTATCAATTAAGAGATTTAGTCTTCCA 4629
 QY 4596 GGTAAACCAACACCGGTGTACAGTGTCTCATCTACT - GCATCAATTTTACTTTGAC 4653
 Db 4630 TGGCAAGCTTATTAACCTGTACGATCTCTTAATCATATGATTTAATTTATTTTGAC 4689
 QY 4654 AGTGAACCTGTAGCCACCTGAGAGACAC - CCAATGTTCCGTTGGTCTCAGATGTACCT 4712
 Db 4690 AGTGAACCTGTAGCCACATGAGAAAGCATCTGTGTTTGTGCTGTCAAGATTTATCT 4749
 QY 4713 AGTGTGCGCGGTGTTGTTTATTTTCAATCTGCAATGTCTTCAACACATTAACATG 4772
 Db 4750 GGTGATGTGTGTTTGTGTTGTTTAAATTTTGGCGTTGTGATGATTAATTAATCA 4809
 QY 4773 TTAGAGCCCACTGCCAGCGGTACGATCATCACTACCAAGCTC - - - - TTAATCTC 4827
 Db 4810 GTAGACCAAC - ACCATGAGAGTGTGTACGATCAAGATATCCACAGCTCTTTTATCTC 4868
 QY 4828 TGTAGAGTGAAG - TTTATTCAGTGTGTTTATGAGAA - - - - TATCTGAAACAATTAAT 4881
 Db 4869 TGTATCATGAAAGTTTATTCAGATTTCTTTTATGAGAAATGACCTATTTTGAACAATTAAT 4928
 QY 4929 TTTCTTACAAAGAAAGATGTATAGAAAGTCTCCCTGCAATTAATTT - CCAATGTTTACAT 4987
 Db 4942 TTTTAACTTGACTGTGGGGGTGCTACAGATTAATTAATTAATGCGCTGCTGCTCGG 5001
 QY 4988 TTTTAACTTGAAGTGT - GGAATTTTACAGATTTAATTAATTAATGAGCTCATGCTCGGT 5046
 Db 5002 TGTGTATTACTTGTGTGATGAGTGAAGCGGTGT - - - - CTTAGATTAATGAGTGAAT 5058

Db 5047 TGTGTGTT-AGATATGCTGTAGCTGAAGCCCTGTTGCTTTTAAACACTAGTTGGAAGC 5105
 Oy 5059 CGGAGAGAGATT 5071
 Db 5106 TGTCAATAAAAT 5118

RESULT 4
 AF127798 4025 bp mRNA linear ROD 28-JUN-1999
 LOCUS Rattus norvegicus EH- and SH3-domain containing protein EHSN1 mRNA,
 DEFINITION complete cds.
 AF127798
 ACCESSION AF127798.1 GI:4835852
 VERSION
 KEYWORDS
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 4025)
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T.C.
 TITLE EHSN1/intersectin, a protein that contains EH and SH3 domains and binds to dynamin and SNAP-25. A protein connection between exocytosis and endocytosis?
 J. Biol. Chem. 274 (26), 18446-18454 (1999)
 JOURNAL 99303609
 MEDLINE 10373452
 PUBLISHED 2 (bases 1 to 4025)
 REFERENCE Okamoto, M., Schoch, S. and Sudhof, T.C.
 AUTHORS Direct Submission
 TITLE Submitted (11-FEB-1999) Center for Basic Neuroscience and HHMI,
 JOURNAL Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA

FEATURES
 source Location/Qualifiers
 1..4025 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 41..3694 /product="EH- and SH3-domain containing protein EHSN1"
 /protein_id="AA030271.1"
 /codon_start=1
 /translation="MAQPTTPTFGSLIDIMAITVEERAKHQDQFSLKPSITGTDQA
 /db_xref="GI:4835853"
 /translation="MAQPTTPTFGSLIDIMAITVEERAKHQDQFSLKPSITGTDQA
 RNFPPGSLPQVLAQIMALADMKGRMDQVPSIANKLILKIQGLQPPALPVM
 KOCPAISGAPAPGIGMAQMPILTAVALPVMPSI.PVGMSPGSLIPQVLAQIM
 GAPVIOIPAPAHRAATLPRKSSFSRSGPSQALTKQKOSPVSAAPAAEAVP
 QSRILKTVLENSHDKTMSGHLTGPOARTILMOSLPOQLASINSSADIDODKLT
 EFFILAMHLDIVAMSGQLPVLPEYI.PPSFRVRSGMSVISISSADQRLPEPS
 SEDVOOEKRLPVTEDKKENFERGNLELEKROALLIQOKERLEADOLEMER
 KEREOERKROLELEKOLEKORLEERERKEI.EEREAARELEADOLEMER
 NROELTORNDQGIIVLKARRTLELEALNDKHLQGLKODIRCLATQROE
 ISTSKSRELTAEITTHLOOQLOESQOMGLIIPKQILSLDKLVQONSLSHRSLT
 LKRALEKELARQQLREQLDEVEKTRSLQELIDVNNOLKELRHSKQQLQKRSI
 EAERLKQKEERKSELEKQKEGQORVQERDQVQEHVQOEQORPKPHEDLKR
 EDSVKKKEAEERAKSEVQDQKSLFPHQEPAPKPAQAMPTEKEGLTISAQESLKV
 YRALYSPESRSHDEITIQGDIVMWKGVDSOTGEQMGSGPKGTGMFPANVA
 EKIPENEIPTPAKPVDTLTSAPAKLALRETPAPLPVSSSEPTTPNNMAWPSSTWPS
 STNEKETNNMTMAAQESLITPSAQQLRQSAFTPATATGSPSPVAGQGEKVGLO
 AALYPMRAKKNHNLNFKSDVITVLEQDDMMFEGVQKMFPSKTVKLISGVK
 STSIDGPTPEASLSLRVASPAKPAIPEEFVAVTYESSHGDLTQOQGHVIVTK
 KQGDWWTGIVGEGTGVPSNVYKRLKDSGSGTAGTLEKKEPEIAQVIASTATGPE
 QTLAPGOLILIRKKNPGGMEGVELQARKKROIEMFPANVYKLSPGSKITPTEL
 KTAVQPAVCQVIGMYDYTAQNQNDLAFSGQIINVLSKEDPDMMKGEVSGQGLPSPN
 YKLTDDMPSSQ"

BASE COUNT 1147 a 1009 c 1123 g 746 t
 ORIGIN

Query Match 69.2%; Score 3517.2; DB 10; Length 4025;
 Best Local Similarity 93.4%; Pred. No. 0;
 Matches 3734; Conservative 0; Mismatches 243; Indels 21; Gaps 5;
 Oy 219 GAGATGAGCGCTGCAATGAGAGTGAAGTGAATAGAACCATGCTGCTGATTTCCACACC 278

Db 1 GAGATGAGCGCTGCAATGAGAGTGAAGTGAATAGAACCATGCTGCTGATTTCCACACC 60
 Oy 279 TTTGCGTGTAGAGCTGTGATGTCTGGCCATATCTGTGAGAAAGGCCAAGCATGCA 338
 Db 61 TTTTGTGGAGAGCTGTGACATCTGGGCATATACGTGAGAGAAAGGCCAAGCATGCA 120
 Oy 339 GCAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATCTGTGATCAGCGAGAACTT 398
 Db 121 GCAGTTCAGAGTTTGAAGCCGATATCTGTGATCAGCGAGAACTT 180
 Oy 399 TTTTTCGATCTGGGTTTACTTCAAGCTGTCTTGAACAATATGGCGCTGACGACAT 458
 Db 181 CTTTTCGATCTGGGTTTACTTCAAGCTGTCTTGAACAATATGGCGCTGACGACAT 240
 Oy 459 GAATAGAGTGAAGATGATGATCAAGTGAATTTCCATAGCCATGAGCTTCAAACT 518
 Db 241 GAATAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 Oy 519 GAAGCTACAAAGATATGAGTCTCCCTCCACACTTCCCTGTGATGAAACAGAACCACT 578
 Db 301 GAAGCTGCAAGATATGAGTCTCCCTCCACACTTCCCTGTGATGAAACAGAACCACT 360
 Oy 579 GCCTATTTCCAGTCCAGCAGATTTGGTATAGAGAGGATTTGATGATGATGATGATGAT 638
 Db 361 GGCATCTCTAGTCCAGCAGGATTTGGTATAGAGAGGATTTGATGATGATGATGATGAT 420
 Oy 639 AGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
 Db 421 AGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 Oy 699 ATCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 758
 Db 481 ATCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Oy 759 GCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 818
 Db 541 GCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 Oy 819 ATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878
 Db 601 ATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 Oy 879 CAGCGCCCTTCAGCAGCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 938
 Db 661 CAGCGCCCTTCAGCAGCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 Oy 939 GTTATTTCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 998
 Db 721 GTTATTTCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 780
 Oy 999 TATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1058
 Db 781 TATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Oy 1059 CATTGATCAAGATGAGAAACTCACTGAGAGAAATTAATCTGATGACCTTAATTTGA 1118
 Db 841 CATTGATCAAGATGAGAAAGCTCACTGAGAGAAATTAATCTGATGACCTTAATTTGA 900
 Oy 1119 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1178
 Db 901 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 Oy 1179 CTTGAGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1238
 Db 961 CTTTGAAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Oy 1239 GAGGCTGCTGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1298
 Db 1021 GGGGCTGCTGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 Oy 1299 GACATTTGAAGATGAAGAGCGGAGAACTTTCGAGCGGAGCTGTGAGCTGGAAGCG 1358

Db 1081 GACATTGAAATPAGAAAGCGGAGAACTTCAGCGAGCAACCTGAGCTGAGAAAGC 1140
Qy 1359 CCGGCAAGCCCTCTTGAGCAGCAGCGCAAAAGAGAGAGCGGTTGCTCAGCTGAGGC 1418
Db 1141 CAGGCAAGCCCTCTTGAGCAGCAGCGCAAAAGAGAGAGCGGTTGCTCAGCTGAGGC 1200
Qy 1419 CGCCGAGCAGAGAGAAAGAGCGGAGCGCCAGAGAGAGAGAGCGGAGCGGAGCTGGA 1478
Db 1201 CGCGGAGCAAGAGAGAAAGCGGAGCGCCAGAGAACTAGAGAGAGAGAGAGAGCTTAA 1260
Qy 1479 GCTGGAAGCAGCTGAGAGAGAGCGGAGCTGAGAGCGGAGCGGAGAGAGAGAGAG 1538
Db 1261 GCTGGAAGAGAGCTGAGAGAGAGCGGAGCGCTGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1539 GAAAGAGATCGAGAGCGCGAGCGCAAAACCGGAACTGAGAAAGCGAGCAACTTGA 1598
Db 1321 GAAAGAGATCGAGAGCGGAGCGCAAAACCGGAGCTGAGAGCGGAGAGAGAGCTTGA 1380
Qy 1599 ATGGGAGCGGAGCGGAGAGAGAGAGCTCGAGATCAGAGAGAGAGAGAGAGAGAG 1658
Db 1381 ATGGGAGCGGAGCGGAGAGAGAGAGCTCGAGATCAGAGAGAGAGAGAGAGAGAG 1440
Qy 1659 CGTGTCTTGAAGCAGAGAGAGAGAGCTTGAAGTTGAGTTGAGTTGAGTTGAGTTG 1718
Db 1441 CGTGTCTTGAAGCAGAGAGAGAGAGCTTGAAGTTGAGTTGAGTTGAGTTGAGTTG 1500
Qy 1719 AAGGATCAGCTGAGAGAGAGAGAGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG 1778
Db 1501 AAGGATCAGCTGAGAGAGAGAGAGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG 1560
Qy 1779 AGAATTTGAGAGCAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAGAG 1838
Db 1561 AGAATTTGAGAGCAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAGAG 1620
Qy 1839 GCAAGCTTGAAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAGAG 1898
Db 1621 GCAAGCTTGAAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAGAG 1680
Qy 1899 CAGGAGCAGTTAAAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 1958
Db 1681 CAGGAGCAGTTAAAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 1740
Qy 1959 CAAAGAGCTTGAAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 2018
Db 1741 CAAAGAGCTTGAAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 1800
Qy 2019 GGTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 2078
Db 1801 GGTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 1860
Qy 2079 GGAATCTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2138
Db 1861 GGAATCTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 1920
Qy 2139 GCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2198
Db 1921 GAGGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 1980
Qy 2199 CGCTCAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2258
Db 1981 AGGTGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2040
Qy 2259 GCAAGCAGCGCCCGGAGAAAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2318
Db 2041 GCAAGCAGCGCCCGGAGAAAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2100
Qy 2319 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 2378
Db 2101 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAGAG 2160
Qy 2379 CCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2438
Db 2161 CCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGTTGAGAGAGTTGAG 2217

Qy 2439 AGCCCGCTTAACTTTCTGACAGAGAGTTGAAAAGTGTATATTAACGAGCGCTGTA 2498
Db 2218 AGGTCCGCTTAACTTTCTGACAGAGAGTTGAAAAGTGTATATTAACGAGCGCTGTA 2277
Qy 2499 CCCCTTTGAATCCAGAGAGTCAAGTATGATCAGATCAGATCAGATCAGATCAGATCAG 2555
Db 2278 CCCCTTTGAATCAGAGAGTCAAGTATGATCAGATCAGATCAGATCAGATCAGATCAG 2337
Qy 2556 -----GGTGAATGAAAAGCAGAGCTGAGAGAGAGAGAGAGTTGAGAGAGAGT 2603
Db 2338 TAAAGGAGATGGGTGATGAAAAGCAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAG 2397
Qy 2604 GAAAGGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2663
Db 2398 GAAAGGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2457
Qy 2664 TCCCACTCCAGCAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAG 2723
Db 2458 TCCCACTCCAGCAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAG 2517
Qy 2724 GCGTGAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2783
Db 2518 GCGTGAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2577
Qy 2784 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2843
Db 2578 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2637
Qy 2844 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2903
Db 2638 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2697
Qy 2904 GAGATCAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2963
Db 2698 GAGATCAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2757
Qy 2964 GAGTGAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3023
Db 2758 GAGTGAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2817
Qy 3024 CAACCACTTAAATTTTAAAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3083
Db 2818 CAACCACTTAAATTTTAAAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2877
Qy 3084 GTGGTTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3143
Db 2878 GTGGTTGAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2937
Qy 3144 TTCAAGGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3203
Db 2938 TTCAAGGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 2997
Qy 3204 TCTTAAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3263
Db 2998 TCTTAAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3057
Qy 3264 CATGATCAGATCAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAG 3323
Db 3058 TATGTATCAGATCAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAG 3117
Qy 3324 TGTGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAG 3383
Db 3118 TGTGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAG 3177
Qy 3384 CTTCCCTTCTAATCTATGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3443
Db 3178 CTTCCCTTCTAATCTATGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3297
Qy 3444 AGGAGTTTGAAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3503
Db 3238 AGGAGTTTGAAGAGAGAGAGAGAGAGAGAGAGTTGAGAGAGAGAGAGAGTTGAGAG 3297


```

OY 3504 TCCGAACAACATCAGCCCTGGCTCTCTGGGAGATGATTTCTGATCCGGAAGAAAGAACCCAG 3563
Db 3298 TCTTGAACAGCTCAGCTGGCTCTCTGGGAGATGATTTCTGATCCGGAAGAAAGAACCCAG 3357
OY 3564 TGGATGCTGGAGAGAGAACTGCAAGCTGAGGAGAAAGCCGAGATGGGTGTTTC 3623
Db 3358 TGGATGCTGGAGAGAGAACTGCAAGCTGAGGAGAAAGCCGAGATGGGTGTTTC 3417
OY 3624 AGCAATTTATGTAAGCTTGAAGCCCGGAACCAAGCAAAATCAGCCGAGATGCTACC 3683
Db 3418 AGCAATTTATGTAAGCTTGAAGCCCGGAACCAAGCAAAATCAGCCGAGATGCTACC 3477
OY 3684 CAAGACCGCAGTCCAGCAGCTGTCAGAGTATGCGGATGATGATTAACCCGCCA 3743
Db 3478 CAAGACCGCAGTCCAGCAGCTGTCAGAGTATGCGGATGATGATTAACCCGCCA 3537
OY 3744 GAAAGTACGAACTGACCTTCAGCAAAAGCCGAGATGATTAAGTCTTCAAGAGAGA 3803
Db 3538 GAAAGTACGAACTGACCTTCAGCAAAAGCCGAGATGATTAAGTCTTCAAGAGAGA 3597
OY 3804 CCGGAGCTGGTGAAGAGAGATGATGATGAGGAGTGGGCTCTCCCAATCCATATGT 3863
Db 3598 CCGGAGCTGGTGAAGAGAGATGATGATGAGGAGTGGGCTCTCCCAATCCATATGT 3657
OY 3864 AAGCTGACCAAGATGAGACCCAGCAGCAATGATATGTTGTCATCCGCC 3923
Db 3658 AAGCTGACCAAGATGAGACCCAGCAGCAATGATATGTTGTCAT -CCCCC 3716
OY 3924 TCAAGCTTGAAGTCTCAAAAGAGACCCATATCCATATCACTGCCAGAGGATGATG 3983
Db 3717 TCAAGCTTGAAGTCTCAAAAGAGACCCATATCCATATCACTGCCAGAGGATGATG 3776
OY 3984 GAGATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4043
Db 3777 GAGATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3836
OY 4044 GAACTACTGACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4103
Db 3837 GAACTACTGACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3896
OY 4104 AGTCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4163
Db 3897 AGTCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3955
OY 4164 GGGGCTTCTAGTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4201
Db 3956 -GTGCTTCTAGTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3992

```

```

RESULT 5
AF064243 5287 bp mRNA linear PRI 21-NOV-1998
LOCUS AF064243
DEFINITION Homo sapiens interseccion short form mRNA, complete cds.
ACCESSION AF064243
VERSION AF064243.1 GI:3859852
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
            Antonarakis, S.E.
TITLE Two isoforms of a human interseccion (ITSN) protein are produced by
            brain-specific alternative splicing in a stop codon
JOURNAL Genomics 53 (3), 369-376 (1998)
MEDLINE 99017974
PUBMED 979604
REFERENCE
AUTHORS Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
            Antonarakis, S.E.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue

```

```

FEATURES
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="21"
        /map="21q22.1-q22.2"
        /lissue_type="brain"
        /dev_stage="fetus"
        107..3769
    CDS
        /codon_start=1
        /product="interseccion short form"
        /protein_id="AACT8610.1"
        /db_xref="GI:3859853"
        /translation="MAQPTPFGSLDIWATVEERAKHQDQFHSKLPISGFTGDOA
            RNFFPGSLDQVLAQWALADNNDMDQVFFSIAMKLIKLOGYDLPALPVM
            KOQVLAISSAPFGMGSIASMPRTVAIPVPMGSIIVGMSPTLVSVPLAIPVLA
            GAPVPIQPLPAFPAHPTATPKSSPSRSGSLNKLKQASFDVASIPVAVMVP
            OSSRLKRLQFNSHDKTMSGHLTPQARTILMOSSLPQKQASINLSDIDQDKLTA
            EEFILAMHLIDVAMSGQPLPPLPPEYIIPSPFRVSGSGISVISTSDVQRLPEEP
            LEDEOQLEKKLPVTEDKKRNFENRGNLEKRRQALIEQKREORLEAOLERAQE
            RKREEROERKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            RNRQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            EISTNSKRLRIATHTHQQQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            TLKRALEKRLAQLHLDQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            MEERLKQKQERKILEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            KRESEVKKQKQERKILEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            KVVYRALYFESRSHDEITIPGDIIVYVSWESVDSQGEPMGLGELGKGTGMPA
            NYAEKIPENVEFPAVQVPTDSTAPAPKALRETPALVATSSPTPNMADFFST
            WPTSTNKEPFTDMDMAAPSLTVSAGQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ
            GLQQLALYFWRKQDNLNKNQDVTIVLEQDQGEFLAMTYSESQDGLTFQGDVIL
            IRKSTMSDSSSPSLKRVASPAKPVVSGEFLAMTYSESQDGLTFQGDVIL
            VTKQDGLMWTGTVQDKAGVPSNYVRLKQSESGTGKTSLEKQLEKQLEKQLEKQLEKQ
            GPBQLTLAPQOLIRKKNPGWMEGLQARKKROIGMPVAVVXLLINQSGKIPPT
            EPESTLAIVCOVIGYDVTQAQNDDELAFNKGQIINVLEKEDPDMWKQGVNGVLF
            PSNYVKTITMDP89Q"
        167..406
        /note="encodes EH domain"
        /note="encodes EH domain"
        767..1936
        /note="encodes EH domain"
        /note="encodes EH domain"
        2324..2524
        /note="encodes SH3 domain"
        /note="encodes SH3 domain"
        2843..3019
        /note="encodes SH3 domain"
        /note="encodes SH3 domain"
        3110..3286
        /note="encodes SH3 domain"
        /note="encodes SH3 domain"
        3326..3520
        /note="encodes SH3 domain"
        /note="encodes SH3 domain"
        3569..3748
        /note="encodes SH3 domain"
        /note="encodes SH3 domain"
        1635 a 1169 c 1227 g 1256 t
    BASE COUNT
    ORIGIN
        Query Match 67.7%; Score 3440.2; DB 9; Length 5287;
        Best Local Similarity 83.8%; Pred. No. 0;
        Matches 4154; Conservative 0; Mismatches 743; Indels 58; Gaps 21;

```

OY	405	CCATCTCTGGGTTACCTCGACGCTGCTTACGACAAATATGGGGCTGAGGGGACATGAAATTA	464
Db	253	TCAAATCTGGGTTACCTCGACCTGTTTTTACGACAGATATGGGCACTAGCTGCATGAAATTA	312
OY	465	CGATGGAAGGATGATCAAGTGAATTTTCCATAGCCATGAACTTATCAAACTGAAGCT	524
Db	313	TGATGGAAGGATGATCAAGTGGAGTTTTCCATATGCTATGAACTTATCAAACTGAAGCT	372
OY	525	ACAAAGGATATAGCTCCCCCTCCACACTTCCCCCGTGCATATGAAACACAAACGATGGCTAT	584
Db	373	ACAAAGGATATAGCTACCTCTGCACTTCCCCGTGCATATGAAACACAAACGATGGCTAT	432
OY	585	TTCCAGTCCACGACATTTGGTATAGGAGGATTTGTCAGCATCCACACGACGACTGT	644
Db	433	TTCTAGGCCACCACTATTGGTATGGGAGGATATGCCAGCATCCACACGCTTACAGCTGT	492
OY	645	TGCTCTCTGTGCCAATGGGCTCCATTTCCAGTTGTTGGAATGTCTTCCACTTATAGTATCTTC	704
Db	493	TGCTCTCAATGTCCAAATGGGATCCATTTCCAGTTGTTGGAATGTCTTCCAAACCTATGATCTTTC	552
OY	705	TGTCCCTCCAGACGAGGCTCCCTCCCTGGCTAAAGGGGAGCTCCCTCCGATACAGGCTCT	764
Db	553	TGTTTCCACAGACGCTGTGTCCCCCTCTGTGTAAGGGGCTCCCCCTGTATATCAACTCTT	612
OY	765	GCTTCGTTTTGCGATCTCTGACGCGACATGGCCAAAGAGTCTTCTCTTACGACAGATCTGG	824
Db	613	GCTCTCATTTTCTCATCTCTGACGCGCATTTGCCAAAGAGTCTTCTCTTATGATAGATCTGG	672
OY	825	TTCCAGGTCACAAATTAAACATAAGTTAAAGGAGGCATCATTTGATGTGGCCAGCGC	884
Db	673	TTCCAGGTCACAACTTAAACATAATTTACAAAGGACACATCATTTGATGTGGCCAGCT	732
OY	885	CCCTTCACAGACAGATGGGCTGTGCTCAGTCATCAAGGCTGTAATAATACAGGAGTTATT	944
Db	733	CCCAACAGTGGAGAGTGGCTGTGTTCTCTCAGTATCAAGGCTGTAATAATACAGGAGTTATT	792
OY	945	CAACAGCCACGACAAAACTATGATGAGACACTTAAACAGGTTCCCCAGGAAAGAACTATTTCT	1004
Db	793	CAATGTATATACAAAACTATGATGAGACACTTAAACAGGTTCCCAAGGAAAGAACTATTTCT	852
OY	1005	CATGCATTCAGATTTACCCCCAGGCTCAGCTGAGCTTCAATATGGAATCTTTTGACATTGA	1064
Db	853	TATGAGTCAGATTTACCAACAGGCTCAGCTGAGCTTCAATATGGAATCTTTTGACATTGA	912
OY	1065	TCAAATGGAATACTCAGTCAGAGAAATTTATCTTAGTATGCACTTAATTGATGTTGC	1124
Db	913	TCAAATGGAATACTTACAGCAGAGGAATTTATCTTGGCAATCACTCATTTGATGTTAGC	972
OY	1125	CATGCTGTGTCAGCCATGCGCCGCTGCTGCGCTCCAGAAATCATCCCTCTTCTCTGAG	1184
Db	973	TATGCTGTGTCAGCCATGCGCACTGTCTGCTCTCCAGAAATACATTTCACTTTCTTTTAC	1032
OY	1185	AAGAGTTGCTCCGCGCAGTGGGATGTCCGTATAGCTCTTCTCTGTGGAATCAAGGCT	1244
Db	1033	AAGAGTTGATCTGGCAGTGGTATATCTGTCAATAGCTCAACATCTGTATGATCAGAGCT	1092
OY	1245	GCTTGAAGGAGCGCTGCTCAGAGGATGAGCGAGCGC---CAGAAAGAAATCGCTGTAC	1304
Db	1093	ACCCAGGGAACCAAGTTTATGAAATGGAACAAACAATTTGAAAGAAATTAATCTGTATAC	1152
OY	1302	ATTGGAATATAGAACGCGGGGAATCTTCAGCGAGGAGGATGTGAGCTGGAGAAGCGCG	1364
Db	1153	GTTTGAATATAAAAAGCGGGGAATCTTTGAACGTGGCACTGGGAATCTGGAAACGAAG	1212
OY	1362	CCAAAGCTCTTTTGAAGCAGCGCCAAAGACAGAGCGGTTGGCTCAGCTGAGCGCG	1424
Db	1213	GCAACCTCTCTTGGAAACGACGCCAAAGAGACAGAGCGGCTTGGCCCACTGGAAGGGCG	1272
OY	1422	CGAGCAGAGAGGAAAGAGCGGGAGCGCCAGAGACAGAGGAGCCAAACGGGAGTGGAGCT	1484
Db	1273	GGAGCAGAGAGGAAAGAGCGTGAAGCGCCAGAGACAAAGAGCGGAAAAGCAACTGGAACCT	1332
OY	1482	GGAGAGCAGCTGGAAGAGCAGCGGAGCTTGAAGCGGACGAGAGAGAGAGAGAGAA	1544

Db 2413 TAAAGGGAATGGGTGATGAAAGCCAACTGAGAACCCGGCTGGCTTGAGAGAAATT 2472
 Qy 2604 GAAAGGAGACGGGATGTTCTCCCTGCAAACTATGCAAGAAAGATTCCAGAAATAGGT 2663
 Db 2473 AAAAGGAAAGACGGGGTGTCTCCCTGCAAACTATGCAAGAAAGATTCCAGAAATAGGT 2532
 Qy 2664 TCCACCTCCGACCAAAACCAATGACATCTGACATCTGCCCCCTCCCAATGCGCTCT 2723
 Db 2533 TCCCGCTCCGATGAACCAATGATCTGATCTGACATCTGCCCCCTCCCAATGCGCTCT 2592
 Qy 2724 GCGTGAGACCCCTGCTCTCTTGGCAGTGACCTCTTGAGACCTCCCAACCCCAACA 2783
 Db 2593 GCGTGAGACCCCGCCCTTGGCAGTGAATCTTTGAGACCCCTCCACGACCCCTTAATA 2652
 Qy 2784 CTGGGCAAGCTTCAATTCCACGTCGCCCCAGACAGCTTAACGAGAACCCAGAAACGAGCA 2843
 Db 2653 CTGGGCGGACCTTCAAGCTCCACGTCGCCCCACGACAGAAATGAGAAACAGAAACGAGTAA 2712
 Qy 2844 CTGGGATAGCTGGGCGGCTCAGCTCTCTGACCGTACCTAGTGGCTGGCAGTTAGGCA 2903
 Db 2713 CTGGGATGATGGGCGGACCGGACCTCTCTCAGCTTCCAGGCGGCGAGTTAAGGCA 2772
 Qy 2904 GAGATGACCTTTTACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2963
 Db 2773 GAGTCCGCTTACTCTCAGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2832
 Qy 2964 GGGTGAAGAGTGAAGGGGTACAGGCGCAAGCCCTGTATCCCTGAGAGCCAAAAAGA 3023
 Db 2833 GGGTGAAGAGTGAAGGGGTACAGGCGCTATATCTTGGAGAGCCAAAAAGA 2892
 Qy 3024 CAACCACTTAAATTTTAAACAAAGTGAACGTCATCACCCTTCTGGAACAGCAGACATGTG 3083
 Db 2893 CAACCACTTAAATTTTAAACAAAGTGAACGTCATCACCCTTCTGGAACAGCAGACATGTG 2952
 Qy 3084 GTGGTTGGAAGATTCAAGGTGAGAGGGTGGTCCCAAGCTTCAAGTGAATTCAT 3143
 Db 2953 GTGGTTGGAAGATTCAAGGTGAGAGGGTGGTCCCAAGCTTCAAGTGAATTCAT 3012
 Qy 3144 TTCAGGCGCCGTAAGAGAAATCCACAGATCGATCTGAGCCCTTCAAGAAAGTCTGCTAG 3203
 Db 3013 TTCAGGCGCCATTAAGAGAGCTCAAGATCGATCTGAGTCTGAGAGTCTGCTAG 3072
 Qy 3204 TCTAAGAGAGTGGCTTCCCGCGCCGCAAGCCGATTTCCCGGAGAAAGATTATTC 3263
 Db 3073 TCTAAGAGAGTGGCTTCCCGCGCCGCAAGCCGATTTCCCGGAGAAAGATTATTC 3132
 Qy 3264 CATGTACACATACAGAGTCTGAGCAGAGAGATTTAACCTTTCAAGAGGGAGTGTAT 3323
 Db 3133 CATGTACACATACAGAGTCTGAGCAGAGAGATTTAACCTTTCAAGAGGGAGTGTAT 3192
 Qy 3324 TGTGTTTACCAAGAAAGATGCTGAGTGAACGAGAAAGGTCGAGCAAGTCCGAGT 3383
 Db 3193 TGTGTTTACCAAGAAAGATGCTGAGTGAACGAGAAAGGTCGAGCAAGTCCGAGT 3252
 Qy 3384 CTTCCTCTTCTAATATGAGGCTTAAAGATTCAAGGGCTCTGCAACTGCTGGGAAAC 3443
 Db 3253 CTTCCTCTTCTAATATGAGGCTTAAAGATTCAAGGGCTCTGCAACTGCTGGGAAAC 3312
 Qy 3444 AGGAGATTAGAAAAAAACCTGAATGCGCAGGTTATGCTCTCAAGCTGCTAGTCCG 3503
 Db 3313 AGGAGATTAGAAAAAAACCTGAATGCGCAGGTTATGCTCTCAAGCTGCTAGTCCG 3372
 Qy 3504 TCCGCAACATCAACCTGCTCTGAGGAGTGAATCTGATCCGAAAAAAAGAACCCAG 3563
 Db 3373 CCGCGAGCAGCTCACTCTGCGCCCTGGTCAAGTGAATTTTATGATCCGAAAAAAAGAACCCAG 3432
 Qy 3564 TGGATGTTGGAAGAGAACTGCAAGCTGAGGAGAAAGCCGACAGATAGGGTGGTTTC 3623
 Db 3433 TGGATGTTGGAAGAGAGCTGCAAGCTGAGGAGAAAGCCGACAGATAGGGTGGTTTC 3492
 Qy 3624 AGCAAAATTATGTCAAACTTAAAGCCCGGAAACAAAGCAAAATCAACCCCAAGTGAAC 3683
 Db 3493 AGCTAATTATGTAAAGCTTTAAACCTTGGAGCAGCAAAATCACTCCAAAGAGCACCC 3552

Qy 3684 CAAGCCGAGTGCAGCCAGAGTGTGCAAGTATCGGATGTAGCATTAACCCGCCA 3743
 Db 3553 TAAATCAACAGCATTTAGGCGAGTGTGCAAGTATGGATGTAGACATACACCGCA 3612
 Qy 3744 GAACGATACGACATGAGCTTCAAGAAAGGCAAGATCATCAAGCTCTCAACAGAGAG 3803
 Db 3613 GAATGACGATAGCTGGCTTCAACAGAGGCGAATATCAACGCTCTCAACAGAGAG 3672
 Qy 3804 CCGGAGCTGTGGAAGAGAGAGTCAAGTGGGCAAGTGGGCTCTTCCATCAATTAATG 3863
 Db 3673 CCGTCACTGTGTGAAGAGAGAGTCAATGACCAAGTGGGCTCTTCCATCAATTAATG 3732
 Qy 3864 AAAGCTGACCAAGACATGACCCGACGACCAATGATCATATGTTGCCATCCCCC 3923
 Db 3733 GAAGCTGACCAAGACATGACCCGACGACCAATGATCATATGTTGCCATCCCCC 3792
 Qy 3924 TCAGGCTTGAAGTCCCAAGAGACCCATATCCATATCACTGCCCCAGAGAGATAG 3983
 Db 3793 TCAGGCTTGAAGTCCCAAGAGACCCATATCCATATCACTGCCCCAGAGAGATAG 3852
 Qy 3984 GGAGATGAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4043
 Db 3853 GGAGATGAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3912
 Qy 4044 GAATCTGAGAGACAGTATTACCTCAATTTGACCTTGAATGATGATGATGATGATGAT 4103
 Db 3913 GAATCTGAGAGACAGTATTACCTCAATTTGACCTTGAATGATGATGATGATGATGAT 3972
 Qy 4104 AGTCACTGCGTCAGAGAGAGACACA--MATGCAAACTGACAGAGGTGGTGGTCTT 4161
 Db 3973 AGTCACTGCGTCAGAGAGAGACACA--MATGCAAACTGACAGAGGTGGTGGTCTT 4032
 Qy 4162 TTGGGCGCTTCTCTGACACTGACAGTGC--CGGCGCGCTTCAAGAGGGGCTTTCAT 4220
 Db 4033 TTGGGCGCTTCTCTGACACTGACAGTGC--CGGCGCGCTTCAAGAGGGGCTTTCAT 4092
 Qy 4221 A-GTTTAAAGATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAAT 4279
 Db 4093 AGTTTAAATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAAT 4152
 Qy 4280 ACTTCTTGGCTATTTTGGTTTACAAAAACCCCATATCAAGAGTGGTGGTGGTGGTGG 4339
 Db 4153 GACTTCTTGGCTATTTTGGTTTACAAAAACCCCATATCAAGAGTGGTGGTGGTGGTGG 4207
 Qy 4340 ACGATTAAGATGCTGTCGCGGCGTAACTGAGAGTGGTGGTGGTGGTGGTGGTGGTGG 4399
 Db 4208 GTGCTATTTAAATTTTCCAAATGTCATTAATCTGAGACTGATGATGATTTTTCATTT 4267
 Qy 4400 TGTCCAGTGTCCCAACACATTTGTAGTTTGGGCTGTTCC--TCCGTAAGCA 4455
 Db 4268 TGTCCAGTGTACCAACATTAATTTGTGAGTTTGGGCTTTTCCCTTACATAGAAAGT 4327
 Qy 4456 CAGAGAGATGGGTGACCTGTTTGAATAATGATGATGATGATGATGATGATGATGATGAT 4515
 Db 4328 CAGAGAGATGGGTGACCTGTTTGAATAATGATGATGATGATGATGATGATGATGATGAT 4386
 Qy 4516 GGGGTATGCTGTCTGAGCAATCAAGTAC--TGTGCGAGATGACATGATGATGATGAT 4574
 Db 4387 GTGTTTGTCTGTTTGTGTGATCAAGCTGATCTTGTGAGCAATGATGATGATGATGAT 4446
 Qy 4575 AAGAAGTAGCT-CTTCTCAATGCTAAACCAACCAAGTACAGTCTGATCTACT 4633
 Db 4447 CATTAAGAAATAGTTCTTTCCATGAGCAAGATTAATCTTGTGAGATGATGATGATGAT 4506
 Qy 4634 --GCATTCATTTTCTTGAACAGTACCTTTGAGCACTGAGAAAGCAAC--CAATGTTT 4690
 Db 4507 TGCATTTAATTTTATTTTGAACAGTACCTTTGAGCACTGAGAAAGCACTGATGTTT 4566
 Qy 4691 CCGTTGATCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4750
 Db 4567 TTGTTGCTGCTCAATTTATCTGTTGATGTTGTTTGTGTTGTTTGTGTTTGTGTTTGTG 4626

Qy	1059	CATTGATCAATATGATAAAACCTCACTGGCAAGAAATTTATCTTACTTGTGACCTTAATTGA	1118
Db	841	CATTGATTAATATGATAAGTCTCAACCGAAGAAATTCATCTTACGATGACACTGATTTGA	900
Qy	1119	TGTTGCCATGTCTGTGTGACGACACTGCGCCCGCTCTGCTCCAGAAATACATCCCTCTTC	1178
Db	901	TGTTGTCTATGTCTGTGTGACGACACTGCGCCCGCTCTGCTCCAGAAATACATCCCTCTTC	960
Qy	1179	CTTCAGAAAGAGTTGGCTTCGGGAGTGGGATGTCCGTCAATAAGCTCTTCTTGTGTGATCA	1238
Db	961	CTTTAGAAAGAGTTGGCTTCGGGAGTGGGATGTCCGTCAATAAGCTCTTCTGCTGTGAGCCA	1020
Qy	1239	GAGGCTGCTGAGAGAGCGGTGTCAGAGGATGAGCAGCAGCAAGAAAGAACTGCTGT	1298
Db	1021	GCGGCTGCTCCAGAGAGCCATGTCAGAGGATGAGCAGCAGTGGAAAAAGAGCTGCTGT	1080
Qy	1299	GACATTTGAATATAAGAAAGCGGAGAACTTCTGAGCGAGGCTGTGGAGCTGAGAAAGCG	1358
Db	1081	GACATTTGAATATAAGAAAGCGGAGAACTTCTGAGCGAGGCAACTGGAGCTGTGAGAAAGCG	1140
Qy	1359	CCGCCAACCGCTCTTTGAGCAGCAGCGCAAAAGAGAGAGCGGTTGGCTACGCTGAGCGC	1418
Db	1141	CAGGCAAGCGCTCTCTGAGCAGCAGCGCAAGAGCAGAGCGGCTGCTCAGCTGTGAGCGC	1200
Qy	1419	CGCGAGAGAGAGAGAAAGAGCGGAGAGCGCCAGAGAGAGAGGCGCAAGCGCGCACTGSA	1478
Db	1201	CGCGAGCAAGAGAGAGAAAGAGCGGAGAGCGCCAGAGAAAGAGAGCGCAAGACAGCTTGA	1260
Qy	1479	GCTGAGAGAGAGAGCTGAGAGAGCAGCGGAGCTGGAGCGCAGCAGAGAGAGAGAGAG	1538
Db	1261	GCTGAGAGAGAGAGCTGAGAGAAAGAGCGGAGCTGSAAGGCGAGAGAGAGAGAGAG	1320
Qy	1539	GAAGAGATTCGAGAGAGCGCGAGCGCGCAAAAACGGGAACTGSAAGAGCAGCGCAACTTGA	1598
Db	1321	GAAGAGATTCGAGAGAGCGCGAGCGCGCAAAAACGGGAGCTCGAAACGGCAGCGCACTTGA	1380
Qy	1599	ATGGGACGGAAACCGGACACAGAGAACTCTGATTCAGAGAGAAAGAGAGCGAGGGGAC	1658
Db	1381	ATGGGACGGAAACCGGAGACAGAGAACTCTCTGATTCAGAGAGAAAGAGACAGAGGGGAT	1440
Qy	1659	CGTGCTCTGAAAGCAAGAGAGAGAGACTCTGGAGTTTGAGTTAGTAAGCTCGAATGACAA	1718
Db	1441	CGTGCTCTGAAAGCGAGAGAGAGAGACTCTGAGTTTGAATTTAGTAAGCTCGAATGACAA	1500
Qy	1719	AAAGCATCAGCTTGAAGAGAAACCTTCAGAGATATCAGGTGTCGACTGCAGACCCAGAGCA	1778
Db	1501	AAAGCATCAGCTGGAAGAGAAACCTTCAGAGATATCAGGTGTGACTGCAGACCCAGAGCA	1560
Qy	1779	AGAAATTTGAGAGCAGAACAGTCTTAGAGAGTTAAGAAATTTGCTGAATCACCCCACTTACA	1838
Db	1561	AGAAATTTGAGAGCAAAACAGTCTTAGAGAGTTGAGAAATTTGCGGAAATCACCCCACTTACA	1620
Qy	1839	CGAGAGTTGACAGAGATCTCAGCAAAATGTTTGAAGACTTATTCAGAGAAACGATACT	1898
Db	1621	CGAACAATTGACAGAGATCTCAGCAGATGCTTGAAGACTTATTCAGAGAAACGATACT	1680
Qy	1899	CAGTACCAAGTTAAACAAAGTCCAGCAGAAACAGTTTGACATAGAGACTGCTTTTACCTT	1958
Db	1681	CAGTACCAAGTTAAACAAAGTCCAGCAGAAACAGTTTGACATAGAGTTGCTTTTACCTT	1740
Qy	1959	CAAAAAGCGCTTTGAAAGCAAAAGAGAGTGGCCCGGACAGACTCTCGGAGAGAGCTGAGACA	2018
Db	1741	CAAAAAGCGCTTTGAAAGCAAAAGAGAGTGGCCCGGACAGACTTTCAGAGAGAGCTGAGAGA	1800
Qy	2019	GGTGGAGAGAGAGACAGGTCAAAGCTGAGAGAGATTGATTTTCAACAACAGCTGAA	2078
Db	1801	GGTGGAGAAAGAGACAGGTCAAAGCTGAGAGAGATTGATTTTCAACAACAGCTGAA	1860
Qy	2079	GGAACTGAGAGAGATACATAGCAAAACAGCAACTCCAGAAAGAGAGTCCCTGAGAGCAGC	2138
Db	1861	GGAACTGAGAGAGATACATAGTAAACAGCAGCTCCAGAAAGAGAGTCCATCGAGGCGGA	1920
Qy	2139	GGCACTGAAGCAAGAAAGACAGAGAGAGAGAGGCTGTGAGTTTAGAGAAAGCAAAAGAGAA	2198

Db 27 GAGAGAGTGGAGCGCCGACGAGGAGGAGCGTAGCTTGCTCCGTAGTACGCGCGCT 86
Qy 67 CGAAGGAGCATCCCGAGCGGCTCCGAGACGCGCGGAGGACGAGCGAGCGGCG 126
Db 87 CGGAGAGAGATCCCGAGCGGCTCCGAGACGCGCGGAGGAGCGGCGGCG 135
Qy 127 GGGATGGTGTGCGCGGCTGCGAGCTCGGCGCTTCTGCGC-GCGGCGTGCAGGCTGACCTGA 185
Db 136 GGGATGGTGTGCGGCGGCTGCGGCTCCGCTCCGAGCGGCGGCTGAGCGGCACTGA 195
Qy 186 TTTGTGTGAGGGGCGGCGCGGCGACCCCGCGAGATGAGGGCTGATCAGTAAGTGA 245
Db 196 TTTGTCTCGGGCGGCGAGCGGACCGCGCGAGATGAGGGTGTGATTAAGCAAGTAA 255
Qy 246 ACCTAATAGAACCATGGCTCAGTTTCCACACCTTTCGAGGTGAGCTGAGTCTGGGCG 305
Db 256 AAGTAAACAGAACCATGGCTCAGTTTCCACACCTTTCGAGGTGAGCTGAGTCTGGGCG 315
Qy 306 CATTAAGTGTGAGGAGGAGGCGCAAGCATGACAGAGTTCTTACGCTGAAGCGGATAGC 365
Db 316 CATTAAGTGTGAGGAGGAGGAGGCGCAAGCATGATCAGAGTTTCATAGTTTAAAGCAATATC 375
Qy 366 GGGATTTATTACTGTGATCAAGCGAGAACTTTTTCATCTGGGTTACTTACGCGC 425
Db 376 TGGATTCTACTGTGTGATCAAGCTAGAAACTTTTTCATCTGGGTTACTTACGCGC 435
Qy 426 TGTCTTAGCACTAATATGAGGCGCTAGCGGACATGAATTAAGATGAGAGTGAATCAGT 485
Db 436 TGTCTTAGCACTAATATGAGGCGCTAGCGGACATGAATTAAGATGAGAGTGAATCAGT 495
Qy 486 GGAATTTTCCATAGCCATGAGCTTATCAAACTGAACTCAAGATATCAGTCCCTC 545
Db 496 GGAATTTTCCATAGCCATGAGCTTATCAAACTGAACTCAAGATATCAGTCCCTC 555
Qy 546 CACACTTCCCCCTGTCTGTAAGACAGACAGCACTGCTATTTCCAGTGCACCACTTTGG 605
Db 556 TGCACCTTCCCCCTGTCTGTAAGACAGACAGCACTGCTATTTCCAGTGCACCACTTTGG 615
Qy 606 TATAGAGGAGATGCTGATGACAGCACTCAGAGCTGCTGCTGCTGCTCAATGGGCTC 665
Db 616 TATAGAGGAGATGCTGATGACAGCACTCAGAGCTGCTGCTGCTGCTCAATGGGCTC 675
Qy 666 CATTCCAGTTTGGAAATGCTCCACCCCTTAGATCTTCTGCTCCCTCAGACAGAGTGC 725
Db 676 CATTCCAGTTTGGAAATGCTCCACCCCTTAGATCTTCTGCTCCCTCAGACAGAGTGC 735
Qy 726 TCCCTGCTGCTAAGGCGGCTCCTCCGCTATACAGCTCTGCTGCTGCTGCTGCTGCTG 785
Db 736 CCCCCTGCTAAGGCGGCTCCTCCCTGCTATACAACTCTGCTGCTGCTGCTGCTGCTG 795
Qy 786 AGCAGCATGGCCAAAGAGTCTTCTCAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 845
Db 796 AGCAGCATGGCCAAAGAGTCTTCTCAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 855
Qy 846 TAAAGTTACAGAGGACATATCTGATGTCGCGAGCGCCCTTCAGACAGAGATGGGCG 905
Db 856 TAAAGTTACAGAGGACATATCTGATGTCGCGAGCGCCCTTCAGACAGAGATGGGCG 915
Qy 906 TGTGCTCAGTCAATCAAGCTGAAATACAGGAGTTATTCACAGCCACGACAAACTAT 965
Db 916 TGTGCTCAGTCAATCAAGCTGAAATACAGGAGTTATTCACATAGTATGACAAACTAT 975
Qy 966 GAGTGCACATTAACAGTCCCGGAGCAAGATCTCATGTGAATCAAGTTTACCCCA 1025
Db 976 GAGTGCACATTAACAGTCCCGGAGCAAGATCTCATGTGAATCAAGTTTACCCCA 1035
Qy 1026 GAGTGCAGTGCCTCAATATGAAATCTTTCGACATTCAGATGAGAGTGAAGTCACTGC 1085
Db 1036 GAGTGCAGTGCCTCAATATGAAATCTTTCGACATTCAGATGAGAGTGAAGTCACTGC 1095
Qy 1086 AGAAGATTTATCTAGTATGACCTTAATGATGTTGCCATGCTGCTGCTGCTGCTGCTG 1145
Db 1096 AGAAGATTTATCTAGTATGACCTTAATGATGATGATGATGATGATGATGATGATGATG 1155

Qy 1146 GCCGCTCTGCTCCAGAAATCATCTCTCTTCTTCCAGAAAGATTGCTCCGCGACTG 1205
Db 1156 ACTGCTCTGCTCCAGAAATCATCTCTCTTCTTCCAGAAAGATTGCTCCGCGACTG 1215
Qy 1206 GATGCTCGTATTAAGCTCTTCTTCTGATCAGAGGCTGCTGAGAGCGCTGCTGAGA 1265
Db 1216 TATATCTGTCTAATGCTTCAATCTGTAGATCAGAGGCTGCTCAAGAGAACCATTTTGA 1275
Qy 1266 GATGAGCAGAGC---CAGAGAGAACTGCTGTGACATTTGAATTAAGAGCGGGA 1322
Db 1276 AGATGAACAAACAAATTAAGAAAGAAATTAAGTAACTGTTAAGATTAAGAGCGGGA 1335
Qy 1323 GAACTTGCAGGAGCAGTGTGAGCTGAGAGAGCGCCCGCAAGCGCTCTTGGAGCA 1382
Db 1336 GAACTTGCAGGAGCAGTGTGAGCTGAGAGAGCGCCCGCAAGCGCTCTTGGAGCA 1395
Qy 1383 GCGCAAGAGCAGAGCGGCTGCTGAGCTGAGAGCGCGCGGAGAGAGAGAGAGAGCG 1442
Db 1396 GCGCAAGAGCAGAGCGGCTGCTGAGCTGAGAGCGCGCGGAGAGAGAGAGAGAGCG 1455
Qy 1443 GAGCGCAGAGCAGAGCGGCAAGCGGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1502
Db 1456 TGAAGCAGAGAGAGAGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515
Qy 1503 GCGGAGCTGAGAGCGGCAAGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
Db 1516 GCGGAGCTGAGAGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1575
Qy 1563 GCGCAAGAGCAGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1622
Db 1576 TGCAG 1635
Qy 1623 ACTCTGATCAG 1682
Db 1636 ACTCTGATCAG 1695
Qy 1683 GACTCTGAGTTGAGTTAGAGAGCTGTAATGACAAAGAGAGAGAGAGAGAGAGAGAG 1742
Db 1696 GACTCTGAGTTGAGTTAGAGAGCTGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1755
Qy 1743 TCGAGATATCAGGTTGCACTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1802
Db 1756 TCGAGATATCAGGTTGCACTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1815
Qy 1803 TAGAGAGCTAAGATTTGAGTGAATCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1862
Db 1816 TAGAGAGCTAAGATTTGAGTGAATCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875
Qy 1863 AATGCTGAGAGACTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1922
Db 1876 AATGCTGAGAGACTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
Qy 1923 GCAAGACAGTTTGCATAGAGACTCGCTTCTTACCTCAAGAGAGAGAGAGAGAGAGAG 1982
Db 1936 GCAAGACAGTTTGCATAGAGACTCGCTTCTTACCTCAAGAGAGAGAGAGAGAGAGAG 1995
Qy 1983 GCTGCGCGGAGCAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2042
Db 1996 ACTAGCTGCGGAGCAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2055
Qy 2043 GCTGAGAGAGATTTGATTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2102
Db 2056 ACTAGAGAGAGATTTGATTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2115
Qy 2103 ACAGCAACTCCAG 2162
Db 2116 ACAGCAACTCCAG 2175
Qy 2163 GAGGAAGAGCCTGAG 2222
Db 2176 ACAGGAAGATTTATTAAG 2235

QY 2223 GGAAGCAATGCTGAGCATGTGACGACGAG---GAGCAGCCAGCCCCCGGAAACC 2279
 DB 2236 GGAAGCAATGCTGAGCATGTGACGACGAGGACGAGCATCAGAACCAAGAAACT 2295
 QY 2280 CCAAGGAGAGACAGACTGAAGAGGAGAACAGTGTACAGAAAGAGAGCGGAAAGAG 2339
 DB 2296 CCAAGGAGAGAGAACTGAAGAGGAGAGAGTGTCAAAAAGAGAGATGGCAGAGAAA 2355
 QY 2340 AGCCAAAGCCGAAATGCAAGACAGAGAGTGGCTTTTCCATCCGATCAGAGCCAGC 2399
 DB 2356 AGCCAAAGCAGAGACAGCAAGCAGCTGGGTGGCTTTTCCATCAACCAAGAAACAGC 2415
 QY 2400 TAAGTGGCCACCCAGGACCCCTGCTACACAGAGAAAGCCGCTTACCATTTCTGC 2459
 DB 2416 TAAGCAGCTGTCCAGGACCTGTGTCCACTGACAGAAAAGGTCCACTTACCATTTCTGC 2475
 QY 2460 ACAGAGAGTGTAAAGTGTATATTACGAGCGCTGTACCCCTTTGAATCCAGAGTCA 2519
 DB 2476 ACAGGAAATGTAAAGTGTATATTACGAGCGCTGTACCCCTTTGAATCCAGAGTCA 2535
 QY 2520 CGATTGATACCATCCAGCCAGAGATATATGTCAT-----GCTGATGA 2564
 DB 2536 TGATTGAATCTATCTACGACAGAGACATATGCTTAAAGGGGAAATGGGTGATGA 2595
 QY 2565 AAGCCAGACTGAGAGAGCAGAGTGGCTTGGAGAGAGCTGAAGAGGAGACGGGATGTT 2624
 DB 2596 AAGCCAACTGAGAGAACCCGGCTGGCTTGGAGAGAAATTAAGAGAAAGACAGAGGTGTT 2655
 QY 2625 CCCTGCAAACTATGACAGAAAAGATTCCAGAAAATGAGGTTCCCATCCAGCCAAACAGT 2684
 DB 2656 CCCTGCAAACTATGACAGAAAATCCAGAAAATGAGGTTCCCGCTCCAGTGAACAGT 2715
 QY 2685 GACCGATCTGACATGCCCCCTGCCCCCAACTGGCTGCGAGAGACCCCTGCTCTT 2744
 DB 2716 GACTGATTCAACATGTGCCCCCTGCCCCCAACTGGCTGCGAGAGACCCCTGCTCTT 2775
 QY 2745 GCCAGTACCTCTTGTAGGCCCTCAACACCCCAACACTGGGAGACTTCACTTCCAC 2804
 DB 2776 GGCAGTAACTCTTCAAGAGCCCTCAAGACCCCTAACTGGGCGGACTTCACTTCCAC 2835
 QY 2805 GTGGCCAGAGCTCAAGAGAGAACGAGAACGAGCACTGGGATATCTGGGCGGCTCA 2864
 DB 2836 GTGGCCAGAGCTCAAGAGAGAACGAGAACGAGTAACTGGGATATCTGGGCGGCTCA 2895
 QY 2865 GCTTCTCTGACCGTACTGAGCTGAGCCAGTTACGAGAGAGATCAAGCTTAAACCCAGC 2924
 DB 2896 GCTTCTCTGACCGTACTGAGCTGAGCCAGTTACGAGAGAGATCAAGCTTAAACCCAGC 2955
 QY 2925 CAGAGCCATGCTCTCTCCCATCTCCCGTCTGGGCGAGGCTGAAGAGGCT 2984
 DB 2956 CAGAGCCATGCTCTCTCCCGTCTCTGCTGGGCGAGGCTGAAGAGGCT 3015
 QY 2985 ACAAGGCGAAGCCCTGATCCCTGAGAGCCAAAGAACCACTTAAATTTTAACA 3044
 DB 3016 ACAAGCTCAAGCCCTATATCTTGGAGAGCCAAAGAACCACTTAAATTTTAACA 3075
 QY 3045 AAGTGAAGTATCAAGCTTGTGAGACAGCAAGCATGTGGTGTGGAGAAATTTCAAG 3104
 DB 3076 AAGTGAAGTATCAAGCTTGTGAGACAGCAAGCATGTGGTGTGGAGAAATTTCAAG 3135
 QY 3105 TCGAAGAGGTGGTGTCCCAAGTCTTACGTGAAGTCAATTTCAAGGCGCTGAAGAAATC 3164
 DB 3136 TCGAAGAGGTGGTGTCCCAAGTCTTACGTGAAGTCAATTTCAAGGCGCTGAAGAAATC 3195
 QY 3165 CACAAGCATGATGAGCCCTACTGAAAGTCTGCTAGTCTAAAGAGATGGCTTCC 3224
 DB 3196 TACAAGATGATGATGCTGTCTTCAAGAGATCCTGCTAGTCTAAAGAGATGGCTTCC 3255
 QY 3225 GGGCGCCAGACCAAGCTTCCGAGAGAGTTTATTCATGATCAATACAGAGTTT 3284
 DB 3256 AGCAGCCAAAGCCGATGCTTCCGAGAGAGAAATTTATTCATGATCAATACAGAGTTT 3315
 QY 3285 TGAAGCAAGGATTTTAACTTTGACAGCAAGGGAGTGTGATTTGTTTCAACAAGAAATG 3344

DB 3316 TGACCAAGAGATTTAACTTTCAAGAGGAGATGATTTTGTATTCACAGAAAGATGG 3375
 QY 3345 TGACTGTGACAGGGAACGGTGGGCGCAAGATCCGAGTCTTCCCTTACTATGTAG 3404
 DB 3376 TGACTGTGACAGGAACAGTGGGCGCAAGAGCCGAGATCTTCCCTTCAATATGTAG 3435
 QY 3405 GCTTAAAGATTCAAGAGGCTCTTGGAATCTGCTGGGAAACAGGAGTTTAAAGAAAAACC 3464
 DB 3436 GCTTAAAGATTCAAGAGGCTCTTGGAATCTGCTGGGAAACAGGAGTTTAAAGAAAAACC 3495
 QY 3465 TGAATTTGCCAGGTATTTGCTTCTTACGCTGCTACTGTGCTCCGAACTCACTGAGC 3524
 DB 3496 TGAATTTGCCAGGTATTTGCTTCTTACGCTGCTACTGTGCTCCGAGAGTCACTCTGC 3555
 QY 3525 TCTGGGCACTGATTTGTATTCGGAAGAAAGAACCCAGTGTGATGTGGAGAGAACT 3584
 DB 3556 CCTGTGTAGCTGATTTTGTATTCGAAAGAAAGAACCCAGTGTGATGTGGAGAGAGCT 3615
 QY 3585 GCAAGCTCGAGGGAAGAGGCGCAGATAGGCTTTCCAGCAATTTATGTCAACTTCT 3644
 DB 3616 GCAAGCAGTGGGAAAGAGCGCAGATAGGCTGTCTCCAGTATTTATGTAAGCTTCT 3675
 QY 3645 AAGCCCGGAAACAGCAAAATCAACCCCACTGAGCTAACCAAGACGCACTGACAGC 3704
 DB 3676 AAGCCCTGGAGACGAGCAAAATCACTCCACAGAGCCACTTAAGTCAACAGCATTTAGCGC 3735
 QY 3705 AGTGTGCAAGTATCGGAGATGTACATTAACATCCGCCCAAGAGATGACGATAGCTT 3764
 DB 3736 AGTGTGCAAGTATCGGAGATGTACATTAACATCCGCCCAAGAGATGATGAGCTT 3795
 QY 3765 CAGCAAGGCGCAGATCACTCAACGCTCTCAACAGAGAGAACCCGAGCTGTGGAAGAGAG 3824
 DB 3796 CAACAGGCGCAGATCACTCAACGCTCTCAACAGAGAGAACCCGAGCTGTGGAAGAGAG 3855
 QY 3825 AGTCAGTGGCAAGTGGGCTCTTCCATCAATATGTATTAAGCTGACACAGACATGA 3884
 DB 3856 AGTCAATGCAAGATGGGCTCTTCCATCAATATGTATTAAGCTGACACAGACATGA 3915
 QY 3885 CCCCAGCAGCAATG 3899
 DB 3916 CCCCAGCAGCAATG 3930

RESULT 8
 AF064244 7247 bp mRNA linear PRI 21-NOV-1998
 LOCUS
 DEFINITION Homo sapiens interseclin long form mRNA, complete cds.
 ACCESSION AF064244
 VERSION AF064244.1 GI:3859854
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 7247)
 Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
 Antonarakis,S.E.
 TITLE
 Two isoforms of a human interseclin (ITSN) protein are produced by
 brain-specific alternative splicing in a stop codon
 JOURNAL
 Genomics 53 (3), 369-376 (1998)
 MEDLINE
 PUBMED
 99017974
 9799604
 REFERENCE
 2 (bases 1 to 7247)
 Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
 Antonarakis,S.E.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
 Michel-Servet, Geneva 4 CH-1211, Switzerland
 FEATURES
 source
 1. 7247
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

CDS

```

/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="Brain"
/dev_stage="fetus"
107..5272
/codon_start=1
/product="Intersectin long form"
/protein_id="AAC7861.1"
/translation="MAOFPPTPGGSLDIWATIVERAKRHOQFSLKIPISGITGDOA
RHFPGSGLPOLVLAQIMALADNNNDNVDVPSIANKLTKLQGLQSLAPLPM
KOQVPAISSAPFGMGITASMPPLTAVAPVPMGSIPIVMOGSIISVPTAVALPM
GAPVYQPLAPAPAPATLPSKSSFSRSGPSGLTKQKQSFVASVPVAEAVP
GSSRLKRYRLFNHSDKTSGLTGPQARTLMQSSLPQAQLAIVNLSDIDQDGLTA
EFFILAMHLIDVMSGQPLPVLPEYIPPSFRVRSAGSIVSISTVDRLPEPV
LEDEOOLEKLPFTFEDKKRPNRPERGKLEKROALEOORKEOEALQAEVCE
REKREBOERKOLEKOLEKORELEKOREKREKREKREKREKREKREKREKRE
RNRROELNQRKQEDIVLAKKKTLEFELEALNDKKHQLEKQDQIRKUTQRO
ELETNRKSELRLAETHLQOQLQESQQLGLPEKQILNDQKQVQNSLHRSTLV
TLKRALAEKELAROHLDQDEVEKTRSKLQEIIPNNQLEKELHNKQLOKQK
MEARLKQKQERKILEKQKEAQRQAEQERKQMLBHOQEDHQRPRKLEKEL
KEESYKQKQEGEKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
KVVYRALLYPESRSHDEITIOGDIIVMKQGWDSOTGSGMTGGELKKGTFAPA
NYAKIPEVNEVPAPKPVPTDTSAPAPKALBETAPPLAVTSSEPTTPNNKADSSIT
WFTNKEKETNDMDMAAQPSTVPSAGQLKQSAFTPATGSSPSVPVQGGKVE
GLQALVPMRAKQNDNLNPNKNDVITVLEQDMMFQEVQKQKQKQKQKQKQKQKQ
IKRSTMSDGSSESPSLKRVASPAKPVSGSEFIAMVYSSQDGLTQOQGVIL
VTKQDMMWTGTVGKAGVPSNVYLRKQDSESGTAGTSGKKEPELAQVIASTAT
GPEQLTLARQILIRKKNPQGWEGELQARQKQKQKQKQKQKQKQKQKQKQKQKQ
EPKSTALAAVCOVIGMTDYTKQNDDELAFNKQIIVNLKEDPVMKQVNGQGLF
PSNVYVLTDDMDPSQWCSDLHLMLPTEKRRQGYIHELIVTEENVNDLQITEL
FQKPLMESELLEKEVAMIFVMKELIMENIKLALVRKQKQKQKQKQKQKQKQKQ
AOLPHQPYIRFCQSLNGALIQOKTBDAPDFKFKVLEMDPRCKMPSSFLIKP
MORVTRYPILIKNILENTPENHDSHKLHAKLEKELCSQVNEGVREKENSRLMI
QAHVQCEGSLPOLVNSVNCIGRPKRLKGLYAKKQKQKQKQKQKQKQKQKQKQ
PLGSGSTDVYFSPKSNLYQKMYKTPILFNEVYVKLPTDPSGDEPIFHSIDRYVTLK
ASINERTAMVQKIRASELYIETEKREKAYLVRSQATGIEGLMNVNVEGELK
CSHGKSNPYCEVTGSOCHITKTIQDPLNPMKNSNGFIIDLEQVLCITVPERDO
FEPDFLGRTEIRVADIKKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
167..406
/note="encodes EH domain"
/misc_feature
767..1936
/note="encodes EH domain"
/misc_feature
2324..2524
/note="encodes SH3 domain"
/misc_feature
2843..3019
/note="encodes SH3 domain"
/misc_feature
3110..3286
/note="encodes SH3 domain"
/misc_feature
3326..3520
/note="encodes SH3 domain"
/misc_feature
3569..3748
/note="encodes SH3 domain"
/misc_feature
3836..4390
/note="encodes SH3 domain"
/misc_feature
4649..4819
/note="encodes PH domain"
/misc_feature
4895..5143
/note="encodes C2 domain"
BASE COUNT 2156 a 1652 c 1747 g 1692 t
ORIGIN

```

Query Match 57.0%; Score 2898.4; DB 9; Length 7247;
 Best Local Similarity 86.4%; Pred. No. 0;
 Matches 3244; Conservative 0; Mismatches 491; Indels 21; Gaps 3;

QY 165 GCGGCGTGGCGGCTGCACTATTTGTGAGGGGCGGCGCGCACCCGCGGAGATG 224
 Db 13 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72
 QY 225 AGGGTGCATCAGCAAGCTAATAGAACCATGCTCAGTTTCCACACTTTTGG 284
 Db 73 AGGGTGCATCAGCAAGCTAATAGAACCATGCTCAGTTTCCACACTTTTGG 132

QY 285 TGTAGCCTGATGTCTGGGCGATATGTGGAGAAAGGCGCAGATGACGACGTT 344
 Db 133 TGGAGCCTGATATCTGGGCGATATGTAGAGAAAGCGAAGCATATCAGCAGTT 192
 QY 345 CCTTACCTGAGGCGATAGGCGGATTTATTCGCGATCAAGCGGAACTTTT 404
 Db 193 CCATGATTAAGCGCAATATCTGATTCATTCGCGATCAAGCGGAACTTTT 252
 QY 405 CCAATCTGGGTTTACCTGAGCTGCTTATAGCAAAATATGGCGCTAGCGCATGAATA 464
 Db 253 TCAATCTGGGTTTACCTGAGCTGCTTATAGCAAAATATGGCGCTAGCGCATGAATA 312
 QY 465 CGATGAGAGATGATCAAGTGAATTTTCCATAGCATGAAGCTTATCAACTGAAGCT 524
 Db 313 TGATGAGAGATGATCAAGTGAAGTTTCCATAGCATGAAGCTTATCAACTGAAGCT 372
 QY 525 ACAAGATATCAGTCCCTCCGACATCCCTCCGATATGAAACAGAACAGAGGCTAT 584
 Db 373 ACAAGATATCAGTCCCTCCGACATCCCTCCGATATGAAACAGAACAGAGGCTAT 432
 QY 585 TTCCAGTGCACACGATTTGATAGAGGATTTGATAGTGCACACGATCAGCTGT 644
 Db 433 TTCTAGGCGACACGATTTGATAGAGGATTTGATAGTGCACACGATCAGCTGT 492
 QY 645 TGCTCTGTGCAATGGGCTCATTCCAGTTTGAATGCTTCACCTTATGATCTTC 704
 Db 493 TGCTCTGTGCAATGGGCTCATTCCAGTTTGAATGCTTCACCTTATGATCTTC 552
 QY 705 TGCTCTGTGCAATGGGCTCATTCCAGTTTGAATGCTTCACCTTATGATCTTC 764
 Db 553 TGCTCTGTGCAATGGGCTCATTCCAGTTTGAATGCTTCACCTTATGATCTTC 612
 QY 765 GCTTCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
 Db 613 GCTTCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
 QY 825 TCCAGGATCAATTTAACTAAAGTTAAGAGGCAATCATTTGATGCTGCGCAGCGC 884
 Db 673 TCCAGGATCAATTTAACTAAAGTTAAGAGGCAATCATTTGATGCTGCGCAGCGC 732
 QY 885 CCCTCCAGCAGCAGAAATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
 Db 733 CCCTCCAGCAGCAGAAATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792
 QY 945 CAACAGCAGCAGCAGAAATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
 Db 793 CAATGATGATGCAAACTATGATGATGATGATGATGATGATGATGATGATGATGAT 852
 QY 1005 CATGCAATCAAGTTTAACTAAAGTTAAGAGGCAATCATTTGATGCTGCGCAGCGC 1064
 Db 853 TATGCAATCAAGTTTAACTAAAGTTAAGAGGCAATCATTTGATGCTGCGCAGCGC 912
 QY 1065 TCAAGATGAGAACTCATGCAAGAAATTTATCTGATGATGCACTTAATGATGCTG 1124
 Db 913 TCAAGATGAGAACTCATGCAAGAAATTTATCTGATGATGCACTTAATGATGCTG 972
 QY 1125 CATGCTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1184
 Db 973 TATGCTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1032
 QY 1185 AAGAGTCTGCTCGGCGAGTGTCCGTCATAGCTCTTCTGATGATGAGGCT 1244
 Db 1033 AAGAGTCTGATCTGCGAGTGTGATATCTGATATAGCTCAACATCTGATGATGAGGCT 1092
 QY 1245 GCTTACGAGCGCTGTGATGAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1304
 Db 1093 ACCAGAGGACCAAGTTTAAAGATGAAACCAATTTAAAGAAATTAATCTGTAAC 1152
 QY 1302 ATTGAAGTAAAGAGCGGAGAACTTGAAGCAGGAGCAGTGTGAGCTGAGAGCGCG 1361
 Db 1153 GTTGAAGTAAAGAGCGGAGAACTTGAAGCAGGAGCAGTGTGAGAGCGAGAG 1212

QY	1362	CCAAAGCCGCTTTGGAGCGACGCGCAAAAGAGCAGAGACGGTTTGGCTCACTGTGACCGGC	1423
Db	1213	GCAAGCTTCTCTGGAAACAGCGCGCAAGAGCAGAGACCGCTTGCCCACTGTGACGGGC	1272
QY	1422	CGAGCAGAGAGGAAAGCGCGGAGCGCCAGAGCAGAGAGGCCAAGCGGCAGCTTGAAGCT	1481
Db	1273	GGAGCAGAGAGAGGAAAGAGCGCTTGAGCGCCAGAGGCAAGCGCGCAAAAGCAACTGGAACT	1332
QY	1482	GGAGAGAGAGCTTGAGAAAGCAGCGCGGAGCTTGAGCGGCGCAGAGAGAGAGAGAGAGAA	1541
Db	1333	GGAGAGAGAACTTGAGAAAGCAGCGCGGAGCTTGAAACGGCAGAGAGAGAGAGAGAGAA	1332
QY	1542	GGAGATCGAGAGGCGCCAGGCGCCAAACCGGAGACTGAAAAGGCAGCGCAACTTGAATG	1601
Db	1393	AGAAATTGAGAGCGAGAGGCTGCAAAACGGAACTTGAAGGCAACGAACTTGAGTG	1452
QY	1602	GGAAAGGAACCGGAGACAGGAACCTCTGAATTCAGAGAAACAGAGAGCAGAGAGGACCGCT	1661
Db	1453	GGAAAGGAATTGGAGGCAAGAACTACTAAATCAAAAGAAACAAAGAACAGAGAGACTAGT	1512
QY	1662	GATCTCGAAGCAAGAGAAAGACTCTGAGAGTTGAGTTAGAAAGCTCTGAATGACAAAA	1721
Db	1513	TGTACTGAAGCAAAAGAAAAGACTTTGGAAATTGGAAATTGAAAGCTCTAAATGATAAAA	1572
QY	1722	GCATCAGCTAAGAGAAAACCTTCAGATATCAGGTGTCCATGCGCAACCCAGAGGCAAGA	1781
Db	1573	GCATCAATTAAGAGGAAAACCTTCAGAGATACAGATGTCCATGTGACCAACCCAAAGGCAAGA	1632
QY	1782	AATTGAGGCAACGAACAAGTCTAGAGGCTTAGAATTTGGTGAATACCCACTTTCACAGA	1841
Db	1633	AATTGAGGCAACAAACAATTTAGAGGTTGAGAAATTCGGAATATACCCATCTACAGCA	1692
QY	1842	GCAGTTGAGGAATCTCAGCAATGCTTGGAGAATTATTCAGAGAAACAGATACTCAG	1901
Db	1693	ACAATTAAGGAATCTCAGCAAAATGCTTGGAGACTTATTCAGAAAACAGATACTCAA	1752
QY	1902	TGACCACTTAAAAACAATGCCAGAGAAACATTTGCATAGAACTCGCTTCTTAACCTCAA	1961
Db	1753	TGACCAATTTAAAAACAAGTTCAGCGAAACAGTTTGCACAGAGATTCACTTGTAACTTAA	1812
QY	1962	AAGAGCTTTGGAGCAAAAGAGAGCTGCGCGGACAGACTCCGGAGACAGCTGAGAGAGT	2021
Db	1813	AAGAGCTTTAAGCAAAAGAACTAGCTCGGCGACACTTCAGAACCAATCGATGAGAGT	1872
QY	2022	GGAGAGAGAGACAGAGCTCAAAAGCTGCGAGAGATTGATGTTTAAACACCACTGAAAGA	2081
Db	1873	GGAGAAAAGAACTAGATCTAAACTACAGGAGATTGATTTTCAATTAATCACTCAAGAA	1932
QY	2082	ACTGAGAGATATCATGAGCAAAACAGCAACTCCGAAGCAGAGGCTCCTCGAGGACGCG	2141
Db	1933	ACTTAAGAGAAATACCAATTAAGCAACAATCCAGAACCAAAAGTCCATGAGAGCTGAACG	1992
QY	2142	ACTGAAGCAGAAAGCAGAGAGAGAGAGACTCTGAGTTAGAGAGCAAAAGAAAGCGC	2201
Db	1993	ACTGAAACAGAAAGACACAGAACCAAAAGATCATATGAATTTGAAAAACAAAAAGAAAGC	2052
QY	2202	TCAGAGAGAGATTTCAGGAABAGGCAACAGCAATGCTCTGAGACATGTGCACAGAGAG---	GA 2258
Db	2053	CCAAAGAGAGACTCAGGAAGAGGCAACAGCACTGTGAGCAATGTCCACAGAGAGAGCA	2112
QY	2259	GCAGCAGCGCCCGGAAACCCACGAGAGAGACAGACTGTAAGAGGGAAGCAAGTGTGAG	2318
Db	2113	GCATCAGAGACCAAGAAAACTCCAGAAAGGAAAAACTGAAAAAGGAGAGAGTGTCA	2172
QY	2319	GAGAGAGAGGCGGAGAGAGAGCCAAAGCCGGAATTCAGACAGACAGAGTCCGCTTTT	2378
Db	2173	AAAGAAAGATGGCAGAGAAAAAGGCAAAACAGAAAGCACAAGACAACTGGGTCCGCTTTT	2232
QY	2379	CCATCCGATCAGGAGCCAGCTTAAGCTGTGGCAACCCAGAGCAACCTGTGCTAACACAGAGAA	2438
Db	2233	CCATTAACAACAGAAACCAAGTTAAAGCCAGCTGTCCAGGCAACCTGTGTCCATCTGCAGAAAA	2292
QY	2439	AGGCCCGCTTACCATTTCTGCACAGAGAGTGTAAAAAGTGTATATTAATCCAGGCGCTGA	2498

Db	2293	AGGTCACCTTACCATTTCTGACAGAGAAAATGTAAAGGGGTATTAACGGGACACTGTA	2352
Oy	2499	CCCTTTGAATCCAGAAGTCAGATGAGATGCACATCCAGCCAGGAATATATGATCAT---	2555
Db	2353	CCCTTTGAATCCAGAAGCCATGATGTAATTCATTCAGCCAGGAGCATATGATATGCT	2412
Oy	2556	-----GTTGATGTAAGAGCAGACTGGAGAGCCAGGATGGCTTGGAGAGAGCT	2603
Db	2413	TAAAGGGGAATGGGTGGATGTAAGAACCAACTGTGAGAACCCGGCTGGCTTGGAGGAGATT	2472
Oy	2604	GAAGAGGAGAAGCGGATGGTTCCTGCAAACTATGCGAGAAAAGATTCCAGAAAATGAGCT	2663
Db	2473	AAAAAGAAAAGACAGGGGTGGTTCCTGCAAACTATGCGAGAAAATCCAGAAAATGAGGT	2532
Oy	2664	TCCACATCCAGCCAAAACCAAGTGCAGATGTGACATCTGGCCCCGCCCCCAACTGGACTCT	2723
Db	2533	TCCGCTTCAGTGAACCAAGTGAATTCAAACATCTGCCCCCTGCCCCAAACTGGCTT	2592
Oy	2724	GCGTAGAACCCCTGCTCTTTGGCAATGACCTTTCTGAGCCCTTCACAAACCCCAACAA	2783
Db	2593	GCGTAGAACCCCGCCCTTTGGCAATGACCTTTCTGAGCCCTTCACAAACCCCAATAA	2652
Oy	2784	CTGGGCGAGACTTCAAGTTCCACGTCGGGCCAGCGCTCAAAACGAGAAACGAGCAAA	2843
Db	2653	CTGGGCGGACTTCAAGCTCCACGTCGGGCCACAGCAAGATGAGAAACGAGAAACGAGTAA	2712
Oy	2844	CTGGGATCACTGGGGCGGCTCAGCTTCTCTGACCCGTACTAGTGTGCGCAGTTACGGCA	2903
Db	2713	CTGGGATGCAATGGGCAAGCCAGCCAGCTCTCTCAACCGTTCCAAAGTGCGGCCAGTTAAAGCA	2772
Oy	2904	GAGATCAAGCTTTTAAACCCAGCCACAGCCATCTGCTCTCCCATTTCCCGTCTTGGGCA	2963
Db	2773	GAGCTCGCCTTTTAACTCCAGCCACGGCCACTGCTCTCCCGTCTCTGTCATAGGCA	2832
Oy	2964	GCGTGAAGAAAGTGGAAAGGGCTTCAACCGCAAGCCCTGATCCCTGGAGAGCAAAAAAGA	3023
Db	2833	GCGTGAAGAAAGTGGAGGGCTTCAAGCTCAAGCCCTATATCTTGGAGAGCAAAAAAGA	2892
Oy	3024	CAACCACTTAATTTTTPAACAAAAGTGAAGCTATCACCGTTCTGGAAACGACAAACATGTG	3083
Db	2893	CAACCACTTAATTTTTPAACAAAATATGATCTATCACCGTCTGGAAACGACAAACATGTG	2952
Oy	3084	GTTGTTTGGAGAAAGTTTCAAGGTGAGAAAGGGTGGTTCCCAAGCTTTCAGTGAAGTCTAT	3143
Db	2953	GTTGTTTGGAGAAAGTTTCAAGGTGAGAAAGGGTGGTTCCCAAGCTTTCAGTGAAGTCTAT	3012
Oy	3144	TTTCAGGCGCCGTAAAGAAATCCACAAGCATCTGATCTGGCCCTTACTGAAGTCTCTGCTAG	3203
Db	3013	TTTCAGGCGCCCATTAAGGAAGTCTACAAGCATGATCTGGTCTTTCAGAGAGCTCTGCTAG	3072
Oy	3204	TCTTAAAGAGTGGCTTCCCCCGCCGCAAGCCAGCATTTCCCGGAGAAAGTTATATGC	3263
Db	3073	TCTTAAACGAGTGGCTTCTTCACAGCACCAAGCCGGTCTTTCGGAGAAAGTTATATGC	3132
Oy	3264	CATGTACACATACGAGAGTCTGAGCAAGAGATTTTAACTTTTCAGCAAGGGAGTGTGAT	3323
Db	3133	CATGTACACATACGAGAGTCTGAGCAAGAGATTTTAACTTTTCAGCAAGGGAGTGTGAT	3192
Oy	3324	TGTGTTTACCAAGAAAGATGTGACTGTGTGACGGGAACGGGTGGCGCAAACTCCGAGT	3383
Db	3193	TTTGTGTTTACCAAGAAAGATGTGACTGTGTGACAGGAACAGTGGCGCGCAAAAGCCGGAGT	3252
Oy	3384	CTTCCCTTCAACATATGTAGAGCTTAAATTCATGAGGGCTCTGGAACTGCTGGGAAAC	3443
Db	3253	CTTCCCTTCAACATATGTGTGGCTTAAATTCATGAGGGCTCTGGAACTGCTGGGAAAC	3312
Oy	3444	AGGAGATTATGAGAAAAAAACCTGAAATTCGCCAGGTTATGTGCTTCTCAAGCTGTACTGG	3503
Db	3313	AGGAGATTATGAGAAAAAAACCTGAAATTCGCCAGGTTATGTGCTCATACCGCCACCGG	3372
Oy	3504	TTCGGAACAACTCAACCTGGCTCTGTGGAGCTGATTTCTGATCCGGAAAAAGAACCCAGG	3563

Db	3373	CCCCAGGACGCTCATCTCTGCCCCCTGGTCACTGATTTTGATTCGGAATAAACCCAGG	3432
Qy	3564	TGGATGTGGGGAAGGAGAACTGCAGGCTTGAGGGAAAAAGCCGAGATVGGGTGTTTCC	3623
Db	3433	TGGATGTGGGGAAGGAGAGCTCCAAGCAGTGGGAAAAAGCCCGAGATVGGCTGGTTCCC	3492
Qy	3624	AGCAATATATGCAACACTTCTTAAGCCCCGGGAACAAGCAAAATCACCCCACTGAGGTACC	3683
Db	3493	AGCTATTATATGTAAGCTTTCAAAACCTTGGGACGAGCAAAATCACTCCAAACAGACACC	3552
Qy	3684	CAAGACCCGAGTGACGACGACGAGCTGTGCAGGTGATTCGGATGTACGATTACCGCCCA	3743
Db	3553	TAAGTCAACAGCATTTAGCGGCGAGTGTGCCAGGTGATTTGGATGTACGATTAACCCCGCA	3612
Qy	3744	GAACGATACGAACACTAGCCTTCAAGCAAAAGCCAGATCATCAACGTCCTTCAACAAGGAGGA	3803
Db	3613	GAATACCATGATGAGCTGGCCCTTCAACAAGGGCCAGATCATCAACGTCCTTCAACAAGGAGGA	3672
Qy	3804	CCCCGACTGTGTGGAAGAAGAGAGTCAGTGGGCAAGTGTGGGCTCTTCCATTCAAATTATGT	3863
Db	3673	CCCTGACTGTGTGGAAGAAGAGTCAATGGAACAAAGTGGGGCTCTTCCATTCCAATTATGT	3732
Qy	3864	AAAGCTGACACAGACATGGAACCCAGCCAGGCAATG	3899
Db	3733	GAAGCTGACACACAGACATGGAACCCAGCCAGGCAATG	3768

LOCUS	AF032118	4103 bp	mRNA	linear	VRT 13-NOV-1999
DEFINITION	Xenopus laevis interseectin mRNA, complete cds.				
ACCESSION	AF032118				
VERSION	AF032118.1	GI:2642624			
KEYWORDS	.				

ORGANISM	Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.	
1 (bases 1 to 4103)	
Yanabhai, M., Hoffman, N. G., Hardison, N. L., McPherson, P. S.,	
REFERENCE	
AUTHORS	

TITLE Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains
JOURNAL U. Biol. Chem. 273 (47), 31401-31407 (1998)

REFERENCE	2 (bases 1 to 4103)
AUTHORS	Hoffman, N.G., Hardison, N.L., Yamabhai, M. and Kay, B.K.
TITLE	Direct Submission
JOURNAL	Submitted (30-Oct-1997) Pharmacol. University of Wisconsin, 1300 University Ave, Madison, WI 53706-1532, USA
FEATURES	Location/Qualifiers
source	1. .4103

CDS 193. .4005

```

//function="involved in endocytosis"
//note="EH domain and SH3 domain containing protein;
similar to intersectin binding proteins Ibp1, encoded by
Genbank Accession Number AF057285, and Ibp2, encoded by
Genbank Accession Number AF057286, and mouse
Rab/Rip, encoded by Genbank Accession Number AF057287"
/codon_start=1
/product="intersectin"
/protein_id="AAC71068.1"
/db_xref="GI: 2642625"
/translation="MAQGFDPFGNLDIMATVEERARHQDFHGLKPTAGYITGDQ
RNFIFSLGIQPVLAQIOWALADNMNDGMDQLSEIANKLTKLKGQGPLSIIPSNN
LKQVAPMAAIVAGGMSGIIVGIPLAIVAPVPMISIVGMSPLTSSVPLPISL
GAPVIOHPAFAHSATLPRKSGRSGASQINIKLOKASFDVPAPIIVEMAVPR
SRLEKROLFNSQDKTMGNLTGPARTILWQSSILPQSQALIMWLSIIDQDGKTLA

```

QY	217	CGGAGATGAGGCGTCGATCAGCAAGGTGAAGCTAATAGAACCATGCTCAGTTTCCACA	276	EEFIAMHLIDVPAWMSGOPLPLIPPEYIIPSEFRRAVSSGSSIMSVSVDORLPPEEPEE
DB	151	CCGCTATAGAGGACCTACTAGCAAGCTAATAGTATACAAACATGCTCAGTTGGAACT	210	EEPNOMLIDVPAWMSGOPLPLIPPEYIIPSEFRRAVSSGSSIMSVSVDORLPPEEPEE
QY	277	CCTTTCGGTGTAGCTTGAGTGTCTGGGCCCATTAAGTGTGAGAGAAAGGCCAAGCATGAC	336	EREODDERKROQDLKOLEKORLEFORBERREKELEREBAKRELEROROLEMEN
DB	211	CCGTTTGGGGGTAAATTGGACATCTGGGCCCAATACGGTGAAGAAAGCATTAACATGAC	270	ERQDLNQNRRDEEDIVLAKAKKTLFELEALNDKHOLEGKJODIRCLTTPQSHI
QY	337	CAGCAGTTCCTTAGCGCTGAAGCCCATTAACGGGATTATTACTGTGTATCAAGCGAGAAC	396	ESTKSLRELIAETIHLQOOLQOESQOQLGMIPEKSOI, IOLKVOQONSIAHRSILTL
DB	271	CAGCAGTTCAGGGGCTCAAGCCAAACACTGATATATACACAGGTGAACAGTAAAGTAAAT	330	ETKRAETLKEIGQOOLRQDLEDEKRETRAKLOEIVDNNKLELRYNKOOFKODEPE
QY	397	TTTTTTTTCCAATCGGGTTACTCAGCGCTGCTTATGACAAATATGCGCGCTAGCGAC	456	TEKIKOLELRKTSJELDKLEKEDRRMLBDQKLMQDVKQDEERYKFODEKEREES
DB	331	TTTTTTCCTTCACTCCGGGCTCCCCCAAGCTGTGCTGCGACAGATATGGGCGCTGTAC	390	VOKEVEKPEBIOEKPRKPHOPREPKLGQPLMMNTEAPLITINGDYKVVYRRAL
QY	457	ATGAATAAAGATGAGAGATGATCAAGTGAATTTTCCATAGCCATGAAGCTTAATCAA	516	YTPARSHDETTLLEPGDIIVDWSQTESPGMLGSELGKGTGWPFAANTAEKMBSEPS
DB	391	ATGAACAATGATGAGAGATGGAACCACTGAGGTTCTCCATAGCTATGAATAATTAACAAG	450	YTKPAABETLAPLTVHAPSPVAPAPAFNTSTNSNMADFSSTWPTNITKVESDNWPT
QY	517	CTGAAGCTACAAAGATATCAGTCCCCCTCCACACTTCCCCCTGTATGAAACAGAACCA	576	MAOPSLTVSAGQHORSAFTPATVYSSPSPLVGGGEKEGEGLOALIPWRKXNDN
DB	451	CTAAAACTACAAAGGTTACCCGTTGGCCCTCTATTTCCTCCATATATTTGAAGCAACCG	510	HLNKKNDVITLVEQODMMWFGEVOGKGSPKRYVLSIGPLKSTSIISTSESPAS
QY	577	GTGGCTAT-----TTCCAGTGCACACACATTTGGTATAGAGGGATGTAGCATGCGCA	630	SLKRVSPAPKPEPAIOGEEYISMTYENBQDLPFOQDILIVIKNQDMMWTVGVKK
DB	511	GTTCGACATCCGCTGCTGCTGCTTGTGCGGGGTTCCGGAGATGATGGAATGTGGGCAATCT	570	TGVPNSVVRPKDBEABGSSGKTSLSKREBIKAVTASYATAPREKOTLTPARQGLILIR
QY	631	CCACTCACAGCTGTGTCTGTGTGCGCAATGGAGCTCCATTCAGTGTGGAAATGTCTCCA	690	KKNGWMBEGLQARGKRKQIGMFPAYVILSLGTAKTSPTPEPKTSLPFTCOVIG
DB	571	CCCCTAGCAGCTGTAGCCCTGTGCGCAATGCATCCATCCAGATGTAGGAATGTCTCCT	630	MYDIAONDELAFASKGVINLVNKEDPDMWKGKLNHVG, FPNYVK, LTTDMP, SQOQ
QY	691	CCCTTAGATATCTTGTGCTCCCTCAGCAGCAAGTCCCTCCCTGGCTTAACGGGGCTCTCTCC	750	FLVLCWPAGGIPATGDRPFTLFFPRDPSLIPNAPLPSVVMIKFRCTAPFCBD

Db 631 CCTCTTGTCTCTGTCC-----CCACAGTTCACCCCTTGTCTAAAGGGGCTCTGCA 684
Qy 751 GTATACAGCCTCGCTGCTTGTGCGATCTGACAGCCCATGGCCAAAGATTTCTCC 810
Db 685 GTATACAGCCTCGCTGCTTGTGCGATCTGACAGCCCATGGCCAAAGATTTCTCC 741
Qy 811 TTGACGAGATCTGGTCCAGGGTCACAATTAACAATTAAAGTTACAGAGGCAATCATTC 870
Db 742 TTGCGCGCTCAGTTGCGGGTCCAGATTAACATTAATCTCAGAAAGCTCAGTCTTC 801
Qy 871 GATGCGCCCGCCCTCCAGACAGGAATGGCTGTGCTCAGTATCAAGCTGAAA 930
Db 802 GATGTCACGCCCCCATTTGTGTGAATGGCTGTGCTCAGTATCAAGCTGAAA 861
Qy 931 TACAGCAGTTATTCACAGCCACGACAAACTATGATGAGACATTAACAGTCCCGCAG 990
Db 862 TACAGCAATTTGTTCAACAGCCAGGACAAACCATGATGGAATTTAAACAGTCTCCAG 921
Qy 991 GCAAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGAAT 1050
Db 922 GCAAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGAAT 981
Qy 1051 CTTTCTGACATTTGATCAAGATGGAATACTCATGACAGAAATTTTACTAGCTATGAC 1110
Db 982 CTTTCAACATTTGATCAAGATGGAATACTCATGACAGAAATTTTACTAGCTATGAC 1041
Qy 1111 CTAAATGATGTTGCTCATGCTGTGCTGACGCACTGCGCCGCTCTGCTCTCAGAAATACATC 1170
Db 1042 TTAAATGATGTTGCTCATGCTGTGCTGACGCACTTCTCTCAATCTGCTCTCAGAGTATAT 1101
Qy 1171 CTTCTCTCTCTCAGAAAGTTGCTGCGGAGTGGAGTGTCCGTCATTAATCTCTTCTCT 1230
Db 1102 CCCCCATCTTTTACAGAGTTTGCATCAGGAGTGGTTATTCATTAATGACTCTGTGTCT 1161
Qy 1231 GTGATCAGAGGCTGCTGAGAGCCGCTGCTCAGAGGATGACGACGCAAGAA 1290
Db 1162 GTTATCAGCGCTTGCGGAGGAAACAGAAAGAGAGAGCCAAACGCGGACAAAG 1221
Qy 1291 CTGCTGTGATTTTGAAGATTAAGAACGGAGAACTTTCAAGCGAGGCTATGAGCTG 1350
Db 1222 CTCCAGTGAATTTGAAGAAAGAAAGGAGAACTTTGAGCGAGAACTGAATTA 1281
Qy 1351 GAGAAGCGCCGCAAGCGCTCTTGAAGCAGGACGCAAGAGCGGATTTGCTCAG 1410
Db 1282 GAAGAGAGAGACAGGACCTTTAGACAGCAGCGCAAGAGCGAGAGATTTGCTCAG 1341
Qy 1411 CTGAGCGCGCGGACGAGAGAGAAAGAGCGGAGCGCCAGAGAGCGAGGCGCAACCG 1470
Db 1342 CTGAAAGGGGCAAGAGAGAGAAAGAAAGAGAGCGGCAAGATCAAGAGCGGAAACCG 1401
Qy 1471 CAGCTGAGCTGAGAGACAGCTGAGAGAGCGGAGCTGAGCGGACGAGAGAG 1530
Db 1402 CAGCAAGACCTTGAAAAACAGCTGAGAGAAAGAGAGAGCTGAGCGGCAACCGGAAAG 1461
Qy 1531 GAGAGAGAGAGAGAGTGAAGCGCGGAGCGGCAAGAAAGCGGAACTGAGAAAGCGAG 1590
Db 1462 GAGGCGAGAAAGAAATCGAAAGAGAGCGGCAAGAAAGAACTGAGAAAGCGAG 1521
Qy 1591 CAACCTGATGAG 1650
Db 1522 CAGTTGAGTGGAG 1581
Qy 1651 GAGGAGACCGTGTCTGAG 1710
Db 1582 GAGGACATTTGTTCTTAAAAAGCAAGAAAGAAAGAAAGCTGAGAGTGGAGAGAGCTTA 1641
Qy 1711 AATGACAAAAAGCATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1770
Db 1642 AATGATAAAAAGCAACATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1701
Qy 1771 CAGAGGCAAGAAATTTGAG 1830
Db 1702 CACGCGCAGAAATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761

Qy 1831 CACTTACAGCAGAGTTCAGAGAACTTCACAAATGCTTGAAGACTTATTCAGAGAA 1890
Db 1762 CACTTACAGCAGAGTTCAGAGAACTTCACAAATGCTTGAAGACTTATTCAGAGAA 1821
Qy 1891 CAGATCTCATGACAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1950
Db 1822 CAGTCTCTTATGATCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881
Qy 1951 CTTACCTCAAAAGAGCTTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2010
Db 1882 CTTACCTCAAAAG 1941
Qy 2011 CTGAGAGAGTGGAG 2070
Db 1942 TTGATGAG 2001
Qy 2071 CAGCTGAAG 2130
Db 2002 CAATTAAG 2061
Qy 2131 GAGCAGCGGATCTGAAG 2190
Db 2062 GAAGCTGAAGAGATCAAGCAG 2121
Qy 2191 AAGAGAGAGCTCAG 2250
Db 2122 AAG 2181
Qy 2251 CAG 2310
Db 2182 CAG 2235
Qy 2311 AGTGCAG 2370
Db 2236 TCCGTTCAAG 2289
Qy 2371 CGGCTTTTCAATCCGATCAG 2430
Db 2290 AAACCGTTCCAGCAACCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2349
Qy 2431 ACAG 2490
Db 2350 ACAG 2406
Qy 2491 GCGCTGTACCCCTTTGAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2550
Db 2407 GCATCTACCTTTTGAATCCAG 2466
Qy 2551 GTCATGTGATGAG 2610
Db 2467 ATCATGTGATGAG 2526
Qy 2611 AAG 2670
Db 2527 AAAAG 2586
Qy 2671 CCAGCAAG 2730
Db 2587 ACAACTAAACAG 2638
Qy 2731 ACCCTGCTCTTTGCGAG 2790
Db 2639 -CTTCCCTGTTGCACTGAG 2697
Qy 2791 GACTTCAGTTCCAGCTGGCCAG 2850
Db 2698 GACTTCAGCTCAAGTGGCCAG 2757
Qy 2851 ACATGGCGAGCTCAAGCTTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2910
Db 2758 ACATGGCGAGCTCAAGCTTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2817

```

QY 2911 GCGTTTACCCAGCAGCAGCAGCTGCTCTCCCATCTCCGCTCTGGGCGAGGGTGA 2970
AK027846
DB 2818 GCTTTTACCCAGCAGCAGCTGCTCTCCCATCTCCGCTCTGGGCGAGGGTGA 2877
QY 2971 AGGTGGAAGGGCTACAGCGCAGCCCTGTATCTCTGGAGAGCCAAAAAGACAAACC 3030
DB 2878 AAGTGAAGGGCTTACAGCAGCAAGGCTGTATCTCTGGAGAGCAAGAACCAAT 2937
QY 3031 TTAATTTTACAAAAGTGCATCACCGTCTGGAACAGCAAGATGTGTGTGT 3090
DB 2938 CTTAATTTTACAAAATATGTATACAGGTTCTTGAACAGAGATATGTGTGT 2997
QY 3091 GGAGAGTTTCAAGTCAAGAGGGTGTCTCCCAAGTCTTACGTAACATCTTCAGG 3150
DB 2998 GGGAGTTTCAAGTCAAGAGGGTGTCTCCCAATCTATGTAAAGCTTATCTCGG 3057
QY 3151 CCGTGAAGAAATCCACAGCATCTGAGCCCTACTGAAAGTCTGTAGTCTAAG 3210
DB 3058 CCACTAAGAAATCTACAGCATCTGAGCCCTACTGAAAGTCTGTAGTCTAAG 3117
QY 3211 AGAGTGGCTTCCCGGCGCCAGCAGCATCTCCGGAAGAGTCTTATGTCATGAC 3270
DB 3118 AGAGTCTTCCAGCAGCATCTTAACTCCCATCAAGAGAAATATCATATCTGAC 3177
QY 3271 ACATACGAGATCTGAGCAAGAGATTTAACTTTCAGCAAGAGATGTGTGTGT 3330
DB 3178 ACATATGAGAGCAAGAGAGAGATTTAACTTTCAGCAAGAGATGTGTGTGT 3237
QY 3331 ACCAAGAAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3390
DB 3238 ATTAAAAAATATGAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3297
QY 3391 TCTAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3450
DB 3298 TCCAACTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3357
QY 3451 TTAGAAAAAAACCTGAAATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3510
DB 3358 CTGGGAAAGAAACAGAAATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3417
QY 3511 CAATCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3570
DB 3418 CAGCTAATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3477
QY 3571 TGGGAAGGAGATCTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3630
DB 3478 TGGGAAGGAGATCTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3537
QY 3631 TATGTCAAACTTCTAAGCTCCCGAAACAGCAAAATCACTGAGCTTCAAGAC 3690
DB 3538 TATGTCAAACTTCTAAGCTCCCGAAACAGCAAAATCACTGAGCTTCAAGAC 3597
QY 3691 GCAATGCAAGCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3750
DB 3598 ACCTCAATGCTCTCTAATCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3657
QY 3751 GACGAATCTAGCTTCAAGCAAGGCAAGATCAAGCTCTTCAAGCAAGGAGAG 3810
DB 3658 GACGAATGCTCTCTAATCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3717
QY 3811 TGGTGAAGAGAGAGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3870
DB 3718 TGGTGAAGAGAGAGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3777
QY 3871 ACCAAGACATGAGACCCAGCAGCAAT 3898
DB 3778 ACCAAGACATGAGACCCAGCAGCAAT 3805

```

```

RESULT 10
AK027846 2131 bp mRNA linear pri 01-AUG-2002
LOCUS Homo sapiens cDNA FLJ14940 fis, clone PLACE1010942, highly similar

```

```

ACCESSION
AK027846
VERSION
AK027846.1 GI:14042823
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens placenta cDNA to mRNA, clone_1lb:PLACE1
clone:PLACE1010942.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K.,
Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Ishii, S., Kawai, Y.,
Saito, K., Yamamoto, J., Makamatsu, A., Nakamura, Y., Nagahara, K.,
Masuno, Y., and Kanehori, K.
NEDO human cDNA sequencing project
TITLE
Unpublished
JOURNAL
2 (bases 1 to 2131)
REFERENCE
1
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction,
5'-3' end one pass sequencing and clone selection: Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
source
1..2131
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1010942"
/tissue="placenta"
/clone_id="PLACE1"
/note="cloning vector: pME18FL3"
BASE COUNT
691 a 448 c 572 g 420 t
ORIGIN
Query Match 27.6%; Score 1404.2; DB 9; Length 2131;
Best Local Similarity 81.0%; Pred. No. 7.1e-312;
Matches 1740; Conservative 0; Mismatches 283; Indels 126; Gaps 4;
QY 7 GAGGAGAGTGGAGGCGGCGGAGGCGGCGGAGTGTGTGTGTGTGTGTGTGTGTGT 66
DB 105 GAGGAGAGTGGAGGCGGCGGAGGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGT 164
QY 67 CGCAAGGAGATCCCGAGCGGCTCCGGAAGCGGCGGAGGAGGCGGCGGCGGCGG 126
DB 165 CGCAAGGAGATCCCGAGCGGCTCCGGAAGCGGCGGAGGAGGCGGCGGCGGCGG 213
QY 127 GGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 185
DB 214 GGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 273
QY 186 TTTGTGTAGGGGGCGGCGGCGGCGGCGGCGGAGATGAGCGGTGATCGAAGGTGA 245
DB 274 TTTGTGTAGGGGGCGGCGGCGGCGGCGGCGGAGATGAGCGGTGATCGAAGGTGA 333
QY 246 ACCTAATGAACCATGAGTCTCAAGTCTTCCACACTTTGGTGTGTGTGTGTGTGTGT 305
DB 334 AAGTAACGAACCATGAGTCTCAAGTCTTCCACACTTTGGTGTGTGTGTGTGTGTGT 393
QY 306 CATTAATCTGTGAAGAAAGGCGCAAGATGACAGAGTCTTGTAGCTGGAAGCGGATAG 365
DB 394 CATTAATCTGTGAAGAAAGGCGCAAGATGACAGAGTCTTGTAGCTGGAAGCGGATAG 453
QY 366 GGGATTTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 425
DB 454 TGGATTTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 513

```

QY 426 TGTCTTACGACAAATATGCGGCTGAGCGCATGAATTAACGATGAAGATGATCACT 485
 Db 514 TGTCTTACGACAAATATGCGGCTGAGCGCATGAATTAACGATGAAGATGATCACT 573
 QY 486 GGAATTTTCATGACGATGAAGCTTATCAAACTGAAGCTACAAAGATATCACTCCCTC 545
 Db 574 GGAATTTTCATGACGATGAAGCTTATCAAACTGAAGCTACAAAGATATCACTCCCTC 633
 QY 546 CACATTTCCCTCTCATGAAACAGAACCCAGTGTATTTCCAGTGCACGCAATTTGG 605
 Db 634 TGACATTTCCCTCTCATGAAACAGAACCCAGTGTATTTCCAGTGCACGCAATTT-- 691
 QY 606 TATAGAGGATGTAGCATGCGCACACTCAGCTGTGCTCTGTGCCAATGGGCTC 665
 Db 692 ----- 691
 QY 666 CATTCAGTTGTGAATGTCTCACCCCTTATCTTCTCTCCACGACGACGCTGCC 725
 Db 692 ----- GCAAGCTGTGCC 702
 QY 726 TCCCTTGGCTAAGGGGCTCTCCGCTATACAGCCCTGCGCTTGGCATCTGC 785
 Db 703 CCCCCTGGCTAAGGGGCTCTCCGCTATACAGCCCTGCGCTTGGCATCTGC 762
 QY 786 AGGCACATGGCCAAAGATTCTTCTTCAAGCATCTGTCCAGGGTCACAATTAACAC 845
 Db 763 AGGCACATGGCCAAAGATTCTTCTTCAAGCATCTGTCCAGGGTCACAATTAACAC 822
 QY 846 TAAGTTACAGAAAGCACATTCATTCATGCGCAGCGCCCTCCAGCAGCAAGAAATGGGC 905
 Db 823 TAAATTAACAAAGGACAGATGATTTGATGCGCAGTGTCCACAGTGCACAGTGCAGTGGGC 882
 QY 906 TGTGCTCATGATCAAGCTGAATATACAGGCACTTATCAACAGCCAGCAAAACTAT 965
 Db 883 TGTTCCTCATGATCAAGCTGAATATACAGGCACTTATCAACAGCTATGACAAACTAT 942
 QY 966 GAGTGAACATTAACAGGCTCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCA 1025
 Db 943 GAGTGAACATTAACAGGCTCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCA 1002
 QY 1026 GGTCTAGCTGGCTTCAATATGSAATCTTTCTGACATTTGATCAAGATGGAATACTGCTGC 1085
 Db 1003 GGTCTAGCTGGCTTCAATATGSAATCTTTCTGACATTTGATCAAGATGGAATACTGCTGC 1062
 QY 1086 AGAAGATTTATCTAGCTATGCACTTAATGATGTTGCCATGCTGTGACCACTGCC 1145
 Db 1063 AGAAGATTTATCTAGCTATGCACTTAATGATGTTGCCATGCTGTGACCACTGCC 1122
 QY 1146 GCCCGTCTGCTTCAGAAATACATCCCTCTTCTTCAAGAAAGTTTCCCTCGGCGAGTGG 1205
 Db 1123 ACCGTCTCTGCTTCAGAAATACATCCCTCTTCTTCAAGAAAGTTTCCCTCGGCGAGTGG 1182
 QY 1206 GATGCTCGATTAAGCTCTTCTTCTGATCAGAGGCTCTCGAGAGCGCTGCGACAG 1265
 Db 1183 TATATCTGTATAGCTCAACATCTGTAGATCAGAGGCTACAGAGGAACTGTTTAA 1242
 QY 1266 GATGAGACAGACG--CAGAGAAAGAACTGCTGTGACATTTGAAAGATTAAGCGGGA 1322
 Db 1243 AGATGAACAACAATTAAGAAAAAATACTGTTAACTTTGAAGATTAAGCGGGA 1302
 QY 1323 GAATCTGACAGGAGGAGCTGTGAGAGGAGCGCCGCAAGCGCTCTGAGAGAGCA 1382
 Db 1303 GAACTTTGAACGTGCAACCTGTGAACCTGAGAAACGAAAGCAAGCTCTCTGAGAGAGCA 1362
 QY 1383 GCGCAAGAGACAGAGCGGTTGGCTCAGCTGAGCGCGCGACAGAGAGAGAAAGAGCG 1442
 Db 1363 GCGCAAGAGACAGAGCGGTTGGCTCAGCTGAGCGCGCGACAGAGAGAGAAAGAGCG 1422
 QY 1443 GGAAGCGCAGAGAGAGAGCGCGCAGCTGTGAGCTGGAAGAGAGAGCTGGAAGAGCA 1502
 Db 1423 TGAAGCGCAGAGAGAGAGCGCGCAAAAGCAACTGTGAGAAAGCAACTGTGAAAGAGCA 1482
 QY 1503 GCGGAGACTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGAGGCG 1562

Db 1483 GCGGAGCTGAGAACAGCAG 1542
 QY 1563 CGCAAAACGGGAACTCTGAAAGGCGAGCGACAACTTTGAATGGGAACCGAACGAGACAGCA 1622
 Db 1543 TGCAAAACGGGAACTCTGAAAGGCGAGCGACAACTTTGAATGGGAACCGAACGAGAGCA 1602
 QY 1623 ACTCTGAAATCAGAGAGAAACAAGAGCAGAGAGGCAACCGTGGTCTGTAAGGCAAGAGAA 1682
 Db 1603 ACTACTAATCAAGAAACAAAGAAACAAGAGGACATAGTGTGTACTGAAACCAAGAA 1662
 QY 1663 GACTGTGAGTTGAGTTAGTAAGCTTGAATGACAAAAACCATCAGCTAGAGAGAAACT 1742
 Db 1663 GACTTTGAAATTTGAATTTAGAGCTTAATGATTAATAAAGCATCACTAGAGAGAAACT 1722
 QY 1743 TCAGAGATTCAGGTGTGACCTGCAACCCAGAGGCAAGAAATTTGAGAGCAAGAAAGTC 1802
 Db 1723 TCAAGATATTCAGATGTGATGATGACACCCAAAGGCAAGAAATTTGAGAGCAAGAAAGTC 1782
 QY 1803 TAGAGAGTAAGAAATTTGCTGAATCACCACCTTACAGACAGCTTGCAGAGATCTCAGCA 1862
 Db 1783 TAGAGAGTTGAGATTTGCGGAATCACCACCTTACAGACAGCTTGCAGAGATCTCAGCA 1842
 QY 1863 AATGCTTGAAGACTTATTTCCAGAGAAACAGATATCTGATGACCACTTAAACAACTCCA 1922
 Db 1843 AATGCTTGAAGACTTATTTCCAGAGAAACAGATATCTGATGACCAATTTAAACAACTCCA 1902
 QY 1923 GCAGAAAGCTTTGATGATGAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGCAAAAGCA 1982
 Db 1903 GCAGAAAGCTTTGATGATGAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGCAAAAGCA 1962
 QY 1983 GCTGCGCCGCGAGAGCTCCGCGAGAGCTTGAAGAGAGTGAAGAGAGAGACCAAGTCAAA 2042
 Db 1963 ACTAGCTGCGAGAGCACTTACAGAGACCAACTGATGATGATGAGAGAGAGAGACCAAGTCAAA 2022
 QY 2043 GCTGCGAGAGATGATGTTTCAACCAACAGCTGAGAGAGAGCTGAGAGATACATAGCA 2102
 Db 2023 ACTAGAGAGATGATGTTTCAACCAACAGCTGAGAGAGAGCTGAGAGATACATAGCA 2082
 QY 2103 ACAGCAACTCCAGAGAGAGAGTCCCTGAGAGGACCGCGACTGAGAGAG 2151
 Db 2083 GCACCACTCCAGAGAGAGAGATCATGAGAGGCTGAGAGAGCTGAGAGAG 2131

RESULT 11
 HSU61166 3241 bp mRNA linear PRI 23-JUL-1996
 LOCUS Human SH3 domain-containing protein SH3P17 mRNA, complete cds.
 DEFINITION U61166
 ACCESSION U61166.1 GI:1438932
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eultheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 3241)
 Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. and
 Kay, B.K.
 Cloning of ligand targets: systematic isolation of SH3
 domain-containing proteins
 Nat. Biotechnol. 14 (6), 741-744 (1996)
 JOURNAL
 MEDLINE
 PUBMED
 98294438
 9630982
 2 (bases 1 to 3241)
 REFERENCES
 Pirozzi, G., McConnell, S.J., Uveges, A. and Fowlkes, D.M.
 Direct Submission
 Submitted (18-JUN-1996) CYTOGEN Corp., 307 College Road East,
 Princeton, NJ 08540, USA
 FEATURES
 Location/Qualifiers
 1..3241
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="bone marrow"

Oy	3969	CCAAAGGAGATGATGGGAGATGACGACCTTATCATGATCTGACACATGATCACCCTCT	4028
Db	1668	CCCAAGGGAGATGATGGGAGATGACGACCTTATCATGATCTGACACATGATCACCCTCT	1722
Oy	4029	GCCCTTCGAG-TAGAAAGACTCACTGACAGACGATTACTCTCATTTTGACCTTAGTTGCAT	4087
Db	1728	GCCCTTCGAGTTAGTAGAAGAACTCACTGCAAAAAGTTTACCCCATTTAACTTAGTTGCAT	1787
Oy	4088	GTGATCGAAATAGTCTGAGTCACTGCGGCGAGAGCGCAAGAACAA-----TTGAGAACTGC	4143
Db	1788	GTTAACCCCAAGTTTGAATTATACCTGGCAAAAATGAAACCAAAATTTTCATPAAACCC	1847
Oy	4144	ACAGGAGTGAGGATCCTTTTGGGCTTTCTTAGTCACTGACTGA-----CCGG	4193
Db	1848	ACGGGAGTGAGGATCCTTTTGGGCTTTCTTAGTCACTGAAATGAATTTATCCCC	1907
Oy	4194	CCCCGCTTACACAGGGGCGCTTCAATAGTTTAAAGTTATTTTAA-----TGT	4244
Db	1908	CACCTTGGCCACAGGGGCGCTTCAAAATTTTAAATTTAAATTTTAAAAAAATTTAT	1967
Oy	4245	GTATTTAGCCTTTTAAATAAATATCTCAATCAATTACTCTTGCCATATTGTGTTT--	4302
Db	1968	TGATTTAACCTTTTAAATACCAAAATATATTAATTAACCTCTGTCTATTTGGGGTTGG	2027
Oy	4303	-ACAAAAACCCCACTATCAAGAGTGCCTGCTCGGAGATTAATAATGCTGTTCCGGG	4361
Db	2028	CAAAAAGACCCCTATTCAAAAGGAAATGCCGCTGTGCTATTAATAAAATAGTTCCAAA	2087
Oy	4362	CGTACCCTTAACTGAGAGCTTGCCTGACCT-----TTGCGCTTGTGCGAGTTCGCAC	4416
Db	2088	ATTTTCCATTAACCGGTAAACCTGAATGTTCTTCTTCAATTTGCTCCGCTGTTACCAAC	2147
Oy	4417	-----CACATGTGTAGTTTGGGCGCTGT-----CC	4442
Db	2148	CTAAATTTCTGCACACTTTGGGGGGCTGTGGTATTTTATTTTTCCTCCCCCTTCAACCC	2207
Oy	4443	CTGCGGTAGACACAGAGAGATGGGTGACCT-----GTTTGAAAATAGTG	4489
Db	2208	CTTCAAAATTTCTCCCAAAAAGAAAGTTTCTCTCTCCCCCCCGCTATATTAACCCA	2267
Oy	4490	TATGTAGACTGAGCTGACTATGAAAGGGGTATAGCTTGTCTGTGACATCACTGCTAC	4549
Db	2268	CACTTAAACTGAAACCCCCCAACCCCCCCTCTAATAAAAAACAATGGTTTTTGGTTC	2327
Oy	4550	TGTGCGCATGTACCATCTGTACCGAAGAAAGTACGCTTCTCTCATAGCTTAAACCCACA	4609
Db	2328	GACCATCTAAAGCATCTCTGCTCTAATAAAATCTTTTTCAGGGCAAAAGCTATTA	2387
Oy	4610	CGGTATCAGTGCCTC--TCATCTACGTCAATCATTTTACTTGC-ACAGTACCTGTGA	4665
Db	2388	CCTTGTAGATGCTCTAATCATATGGCATTTTATTTTATTTTGGCAACGTACCTTGTA	2447
Oy	4666	GCCACCTGAGAGAGAC-CCATGTTCCGTTTGTGCTCAGATGTACTAGTTGTCGGT	4724
Db	2448	GCCACATAGAGAAACACACTGTGTGTTTGTGCGTCTCAGATTTATCTGTGTGAATGGT	2507
Oy	4725	GTTTGTGTTTATTTTTCATCTGCGATGTCTTACACACAAATAACTAGTAAGAACGCCAC	4784
Db	2508	GTTTGTGTTTGGGCTTTTAAATTTTGTGCTGTTTGATGACATTAATACTAGTAGAACAC-AC	2566
Oy	4785	TGCCAGCGGTTTACGATCATCAAGTACCAACCGCTC-----TTAGTCTGTAGCTGAG	4839
Db	2567	CACGTAGGTCGTTACGATCAACGATATCCACAGTCTCTTTTAGTCTGTAGACTGAG	2626
Oy	4840	TTTATTTCACTGTGCTTTTATGAA-----TATCTTGAACAAGTATCTTTTGAACAAG	4893
Db	2627	TTTTATTTCACTGTCTTTTCAATGGAAGACCTATTTTGAACAAGTATTTTCTTGAACAAG	2686
Oy	4894	AAAAGATATATAGAAGTCTCCCTGCATATTAATTTCCAGGCTTACAA--TTTTTAACT	4950
Db	2687	AAAAGATATATAGAAGTCTCCCTGCATATTAATTTCCCAAGTTTAAAAATTTTTTAAATA	2746
Oy	4951	AGACTGTGGGGGTGCTACAGATTAAAT-ATGAATATGGCGCTCTGCTGCTGTGTGTGT	5008

Db 2747 GGACTGCGGAATTTTAAAGATTATATGAAAAGAGCTCAGGTCCTTTGGGGGT 2806

Qy 5009 TACTTGTGCTAGCTGAAGCCGTGTCTTAGATTAGTTGGAATCGG 5061

Db 2807 AAGAAAGCTGTAGGGAAAGCCCTCTTTGTTTTAAACACTAGGTGAAG 2859

RESULT 12			
AF169621			
LOCUS	AF169621	1133 bp	linear
DEFINITION	Esel protein mRNA, partial cds.		
ACCESSION	AF169621		
VERSION	AF169621.1	GI:5731280	
KEYWORDS			
SOURCE	Mus musculus.		
			ROD 12-AUG-1999

Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Mus
1 (bases 1 to 1133)

AUTHORS	TITLE
Isyda, L.O., Kvasna, S.M., Skripkina, I.Y., Anoprienko, O.V., Slavov, D., Tassone, F., Rynditch, A.V. and Gardiner, K.	Mouse homologues of human chromosome 21 genes

REFERENCE
AUTHORS
2 (bases 1 to 1133)
Tsyba, I.O., Kvasha, S.M., Skripkina, I.Y., Anoprienko, O.V.,
Slavov, D., Tassone, F., Ryndtch, A.V. and Gardiner, K.

TITLE Direct Submission
JOURNAL Submitted (15-JUL-1999) Department of Molecular Oncogenetics,
Institute of Molecular Biology and Genetics of National Academy of
Sciences of Ukraine, 150, Zabolotnogo str., Kiev 252627, Ukraine

FEATURES
source

```

/organism="Mus musculus"
/dev_xref="taxon:10090"
/dev_stage="15-day embryo"
<1. .749
CDS

```

```

/note="similar to Homo sapiens interseclin"
/codon_start=3
/product="Hes1 protein"
/protein_id="AAB48848.1"
/db_xref="GI:5731281"
/translation="VRKSTSIDTGPTEPSASLKRVASPAKPAIPGEEFTAMTYESS
EGLGFLPQQQDVIIVTKKDDMWGTGTGDSGVPPSNVYRLKSEGGTAKTGLSK
KPELAIQVIASTAGPOLFLAPGOLLIRKNPGMWEEGLDARGKKRQIGMPAYW
VLLSPGTSKITPTELPKTAVOPACVIMQYVYTAQNDELAFLSKQIINVLNKEDP
DMWKGESGVQGLPPSNVYVLTITDMDSSQ"

```

Query Match	21.8%	Score 1110	DB 10	Length 1133
Best Local Similarity	99.7%	Pred. No. 3.1e-24		
Matches 1133	0	Mismatches 0	Indels 3	Gaps 2

QY 3152 CCGTAAGGAATTCACAGATGTGATATCGGCCCTTCTAAAGTCTCTGTACTTAAGA 32111

Ddb 1 CCGTAAGGAATTCACAGATGTGATATCGGCCCTTCTAAAGTCTCTGTACTTAAGA 60

QY 3212 GAGTGCCTCCCGGCGCCAGCCAGCATTCGCGAAGAAGTTATTGCATGTACA 32171
Db 61 GAGTGCCTCCCGGCGCCAGCCAGCATTCGCGAAGAAGTTATTGCATGTACA 120

Oy 3272 CATTCGAGAGTTCGACGAAGAGATTAACTTTCCAGCAAGGGATGTGATTGCGTTA 3333
|||
D6 121 CATTCGAGAGTTCGACGAAGAGATTAACTTTCCAGCAAGGGATGTGATTGCGTTA 180

Qy 3332 CCAAGAAAGATGCTGACTGCTGACCGGAAACGTTGGCCACAAGTCCGAGTCTTCCCTT 3391
Db 181 CCAAGAAAGATGCTGACTGCTGACCGGAAACGTTGGCCACAAGTCCGAGTCTTCCCTT 240

Qy 3392 CTAACATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACTCTCTGGAAAAACAGGAGTT 3451
Db 241 CTAACATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACTCTCTGGAAAAACAGGAGTT 300

QY 3452 TAGAGAAAAAAGCTGAATTTGCCAGTTATTCCTTCTACGCTGTCTACTGTCCGAAAC 3511
 DB 301 TAGAGAAAAAAGCTGAATTTGCCAGTTATTCCTTCTACGCTGTCTACTGTCCGAAAC 360
 QY 3512 AACTCACCCTGCTCTCTGGGAGCTATTTCTGATCCGAAAAAAGAACCCAGGTGATGGT 3571
 DB 361 AACTCACCCTGCTCTCTGGGAGCTATTTCTGATCCGAAAAAAGAACCCAGGTGATGGT 420
 QY 3572 GGGAGAGAGAACTGAAGCTCCGAGGAAAAAGGCGAGATAGGAGTGTTCAGCAAAAT 3631
 DB 421 GGGAGAGAGAACTGAAGCTCCGAGGAAAAAGGCGAGATAGGAGTGTTCAGCAAAAT 480
 QY 3632 ATGTCAAACTTCTTAAGCCCCCGAAGCAAAATCAACCCAACTGAGTACCAGAACCG 3691
 DB 481 ATGTCAAACTTCTTAAGCCCCCGAAGCAAAATCAACCCAACTGAGTACCAGAACCG 540
 QY 3692 CAGTCAGAGCAGAGATGTGTCAGAGTATGGGATGTACATTAACCCGCCAGAACGATG 3751
 DB 541 CAGTCAGAGCAGAGATGTGTCAGAGTATGGGATGTACATTAACCCGCCAGAACGATG 600
 QY 3752 AGCAATAGGCTTCAGCAAAAGGCGAGATCAACAGTCTCCAAAGAGAGACCCGAGCT 3811
 DB 601 AGCAATAGGCTTCAGCAAAAGGCGAGATCAACAGTCTCCAAAGAGAGACCCGAGCT 660
 QY 3812 GGTGAGAGAGAGAGTCAAGTGGGCAAGTGGGCTCTTCCATCCAAATTAATGAAGCTGA 3871
 DB 661 GGTGAGAGAGAGAGTCAAGTGGGCAAGTGGGCTCTTCCATCCAAATTAATGAAGCTGA 720
 QY 3872 CCACAGACATGACCCCGAGCAGCAATGAATCAATATGTGTTCATCCCCCTCAGGCTT 3931
 DB 721 CCACAGACATGACCCCGAGCAGCAATGAATCAATATGTGTTCATCCCCCTCAGGCTT 780
 QY 3932 GAAAGTCTCAAGAGAGACCACTATCCCATATCACTGCCAGAGGATGATGGAGATGC 3991
 DB 781 GAAAGTCTCAAGAGAGACCACTATCCCATATCACTGCCAGAGGATGATGGAGATGC 840
 QY 3992 AGCCTTGATCATGTGATCTGACAGCATGATCACTACTGCTTGAAGTGAAGAACTCAC 4051
 DB 841 AGCCTTGATCATGTGATCTGACAGCATGATCACTACTGCTTGAAGTGAAGAACTCAC 900
 QY 4052 TGCAGAGCAGTAACTCACTTGAATGATGTGATGATGCAAAATGTCTGATGACTG 4111
 DB 901 TGCAGAGCAGTAACTCACTTGAATGATGTGATGATGCAAAATGTCTGATGACTG 960
 QY 4112 CGTGACAGGAGAGAGAAATTCAGAACTGCAAGGAGTGGTCTTTGGGCTTT 4171
 DB 961 CGTGACAGGAGAGAGAAATTCAGAACTGCAAGGAGTGGTCTTTGGGCTTT 1020
 QY 4172 CTTAGTCACTCAGACCTGACCGGCGCTTTCACAGCGGCGCTTTCATATGTTTAAGAT 4231
 DB 1021 CTTAGTCACTCAGACCTGACCGGCGCTTTCACAGCGGCGCTTTCATATGTTTAAGAT 1079
 QY 4232 TATTTTAAATGTGATTTTATGCTTTTAAATTAATTAATTAATTAATTAATTAATTA 4287
 DB 1080 TATTTTAAATGTGATTTTATGCTTTTAAATTAATTAATTAATTAATTAATTAATTA 1133

RESULT 13
 AC126053/c 230097 bp DNA linear HTG 02-JUL-2002
 LOCUS Mus musculus chromosome UNK clone RP23-280L21, WORKING DRAFT
 DEFINITION Mus musculus chromosome UNK clone RP23-280L21, WORKING DRAFT
 AC126053
 AC126053 1 GI:21672255
 ACCESSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 230097)
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE The sequence of Mus musculus clone

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 230097)
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.edu
 Project Information
 Center project name: M_BA0280L21
 ----- Summary Statistics -----
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 225369 bases at least Q40
 Consensus quality: 226638 bases at least Q30
 Consensus quality: 227672 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1570: contig of 1570 bp in length
 * 1571 1670: gap of unknown length
 * 1671 5006: contig of 3336 bp in length
 * 5007 5106: gap of unknown length
 * 5107 10590: contig of 5484 bp in length
 * 10591 10690: gap of unknown length
 * 10691 16151: contig of 5461 bp in length
 * 16152 16251: gap of unknown length
 * 16252 24144: contig of 7893 bp in length
 * 24145 24244: gap of unknown length
 * 24245 37242: contig of 12998 bp in length
 * 37243 37343: gap of unknown length
 * 37343 50683: contig of 13341 bp in length
 * 50684 50783: gap of unknown length
 * 50784 68844: contig of 18060 bp in length
 * 68844 68943: gap of unknown length
 * 68943 94655: contig of 25712 bp in length
 * 94656 94755: gap of unknown length
 * 94756 136120: contig of 41365 bp in length
 * 136121 136220: gap of unknown length
 * 136221 176981: contig of 40761 bp in length
 * 176982 177081: gap of unknown length
 * 177082 230097: contig of 53016 bp in length.
 Location/Qualifiers
 1. 230097
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-280L21"
 1. 1570
 /note="assembly_name:Contig14"
 1671. 5006
 /note="assembly_name:Contig16"
 5107. 10590
 /note="assembly_name:Contig17"
 10691. 16151
 /note="assembly_name:Contig18"
 16252. 24144
 /note="assembly_name:Contig19"
 24245. 37242

misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature


```

misc_feature /note="assembly_name:Contig20"
37343..50683
/note="assembly_name:Contig21"
misc_feature /note="assembly_name:Contig21"
50784..68843
/note="assembly_name:Contig22"
misc_feature /note="assembly_name:Contig22"
68944..94655
/note="assembly_name:Contig23"
misc_feature /note="assembly_name:Contig23"
94756..136120
/note="assembly_name:Contig24"
misc_feature /note="assembly_name:Contig24"
136221..176981
/note="assembly_name:Contig25"
misc_feature /note="assembly_name:Contig25"
177082..230097
/note="assembly_name:Contig26"
BASE COUNT 60638 a 54726 c 53286 g 60340 t 1107 others
ORIGIN
Query Match 20.4%; Score 1039.4; DB 2; Length 230097;
Best Local Similarity 98.5%; Pred. No. 9.7e-228;
Matches 1145; Conservative 0; Mismatches 6; Indels 12; Gaps 9;
QY 3895 CAATGAATCATATGTTGTCATCCCTCCAGGCTTGAAGAGTCTCAAGAGAGCCACT 3954
Db 178452 CTAGGAATCATATGTTGTCATCCCTCCAGGCTTGAAGAGTCTCAAGAGAGCCACT 178393
QY 3955 ATCCCATATCATGCTCCAGAGGATGATGGAGATGAGAGCTTGAATCATGCTGAG 4014
Db 178392 ATCCCATATCATGCTCCAGAGGATGATGGAGATGAGAGCTTGAATCATGCTGAG 178333
QY 4015 CATGATCCTACTGCTGCTGAGTGAAGAACTCACTGAGAGAGCTTGAATCATGCTGAG 4074
Db 178332 CATGATCCTACTGCTGCTGAGTGAAGAACTCACTGAGAGAGCTTGAATCATGCTGAG 178273
QY 4075 ACCTTGAATGATGATGATGCAAAATGCTGAGTCACTGCTGAGAGAGAGCAAAATG 4134
Db 178272 ACCTTGAATGATGATGATGCAAAATGCTGAGTCACTGCTGAGAGAGAGCAAAATG 178213
QY 4135 CAGAACTGCAAGAGGATGATGCTGCTGAGGCTTCCAGTCACTGAGAGAGAGAGAG 4194
Db 178212 CAGAACTGCAAGAGGATGATGCTGCTGAGGCTTCCAGTCACTGAGAGAGAGAGAG 178154
QY 4195 CCGGCTTCAACAGGAGGCTTCAATGATGATGATGATGATGATGATGATGATGATG 4254
Db 178153 CCGGCTTCAACAGGAGGCTTCAATGATGATGATGATGATGATGATGATGATGATG 178094
QY 4255 CTTTAAATTAATAATTCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 4314
Db 178093 CTTTAAATTAATAATAATTCATCAATCAATCAATCAATCAATCAATCAATCAATCA 178037
QY 4315 ACTATCAAGAGAGTCTGCTGCTGAGAGATGATGATGATGATGATGATGATGATGATG 4374
Db 178036 ACTATCAAGAGAGTCTGCTGCTGAGAGATGATGATGATGATGATGATGATGATGATG 177981
QY 4375 GAGAGCTTGTCTGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4434
Db 177980 GAGAGCTTGTCTGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 177922
QY 4435 GCTGTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4494
Db 177921 GCTGTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177862
QY 4495 AGACTGAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4554
Db 177861 AGACTGAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177802
QY 4555 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4614
Db 177801 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 177743
QY 4615 TACAGTCTCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTG 4674
Db 17742 TACAGTCTCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTG 177663
QY 4675 GGAAGCACCATGTTTCCGTTGCTCAGATGATCTAGTGTGCTGTTGTTGTTGTTT 4734

```

```

Db 177682 GGAAGCACCATGTTTCCGTTGCTCAGATGATCTAGTGTGCTGTTGTTT 177623
QY 4735 TATTTTCAATCTGAGATGCTTCCACACCATTAATCTAGTGAAGAGAGAGAGAGAGAGAGAG 4794
Db 177622 GAGTTTCAATCTGAGATGCTTCCACACCATTAATCTAGTGAAGAGAGAGAGAGAGAGAGAG 177563
QY 4795 GTTACGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4854
Db 177562 GTTACGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 177503
QY 4855 TTTTATGGAATATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4914
Db 177502 TTTTATGGAATATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177443
QY 4915 CTGCAATTAATTTCCAGTGTGTTTAACTAATGATGATGATGATGATGATGATGATGATGATG 4974
Db 177442 CTGCAATTAATTTCCAGTGTGTTTAACTAATGATGATGATGATGATGATGATGATGATGATG 177385
QY 4975 AATATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5034
Db 177384 AATATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 177325
QY 5035 TGTCCTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5097
Db 177324 TGTCCTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 177302

RESULT 14
AF180522 1996 bp mRNA linear PRI 05-SEP-1999
LOCUS Homo sapiens intersecin short form 2 (ITSN) mRNA, partial cds.
DEFINITION AF180522
ACCESSION AF180522.1 GI:5823551
VERSION AF180522.1 GI:5823551
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1996)
Teyba,L.O., Kvashina,S.M., Skripkina,I.Y., Anoprienko,O.V.,
Slavov,D., Tassone,F., Rynditch,A.V. and Gardiner,K.
Mouse homologs of human chromosome 21 genes
Unpublished
JOURNAL 2 (bases 1 to 1996)
Teyba,L.O., Kvashina,S.M., Skripkina,I.Y., Anoprienko,O.V.,
Slavov,D., Tassone,F., Rynditch,A.V. and Gardiner,K.
Direct Submission
TITLE Submitted (25-AUG-1999) Department of Molecular Oncogenetics,
JOURNAL Institute of Molecular Biology and Genetics of National Academy of
Sciences of Ukraine, 150, Zabolotnogo str., Kiev 252627, Ukraine
FEATURES
source
1..1996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="infant"
<1..1996
/gene="ITSN"
<1..566
/gene="ITSN"
/codon_start=3
/product="intersecin short form 2"
/protein_id="AA053183.1"
/db_xref="GI:5823552"
/translation="KSYVKILSGPIRKSTMSGSESPASLKRVAAPAKVSGEE
IAQVIVSYTAAGBEOLTLAPGQILIRKNPRGWMGBLQARKKQIQWFPANVYKL
LSPTSKITPTPEPKSTALAAVCOVIGWYDTPAONDDELAFNKGQIINLVNKEDPDMW
KGEVNGQVGLFSPNIVYKLTITDMP90Q"
BASE COUNT 552 a 431 c 429 g 584 t
ORIGIN

```


Query Match 17.4%; Score 883; DB 9; Length 1996;
 Best Local Similarity 78.8%; Pred. No. 5,5e-192;
 Matches 1286; Conservative 0; Mismatches 310; Indels 37; Gaps 18;

QY 3466 GAAATTTGCCAGGTTATGCTTCTTACGCTGCTACTGTCGCCAGACACTCACCCTGGCT 3525
 Db 132 GAAATTTGCCAGGTTATGCTTCTTACGCTGCTACTGTCGCCAGACACTCACCCTGGCT 191
 QY 3526 CCGGGGAGCTGATTCGTGATCCGGAAAAAGAACCCAGTGGATGGTGGGAGAGAGACTG 3585
 Db 192 CCGGGGAGCTGATTCGTGATCCGGAAAAAGAACCCAGTGGATGGTGGGAGAGAGACTG 251
 QY 3586 CAAGCTCGAGGAAAAAGGCCAGATAGGGTGGTTTCAGCAAAATTATGTCAAATCTTCTA 3645
 Db 252 CAAGCAAGTGGAAAAAGGCCAGATAGGGTGGTTTCAGCAAAATTATGTAAAGCTTCTA 311
 QY 3646 AGCCCTGGGACGAGCAAAATCACTCCACAGACCACTTAAGTCAACACATTAAGCGCA 3705
 Db 312 AGCCCTGGGACGAGCAAAATCACTCCACAGACCACTTAAGTCAACACATTAAGCGCA 371
 QY 3706 GTGTGACAGGTGATCGGATGTAGATTAACCGCCAGACGATAGCAACTAGCTTC 3765
 Db 372 GTGTGACAGGTGATCGGATGTAGATTAACCGCCAGACGATAGCAACTAGCTTC 431
 QY 3766 AGCAAGGCCAGATCACTCAACGTCCTCAACAGAGAGACCCGACTGTGGAAAGAGAA 3825
 Db 432 AACAGAGGCCAGATCACTCAACGTCCTCAACAGAGAGACCCGACTGTGGAAAGAGAA 491
 QY 3826 GTGAGTGGGCAAGTGGGCTCTTCCCATATTATGTAAAGCTGACACAGATGAC 3885
 Db 492 GTGAGTGGGCAAGTGGGCTCTTCCCATATTATGTAAAGCTGACACAGATGAC 551
 QY 3886 CCAGGACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3945
 Db 552 CCAGGACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 611
 QY 3946 AGACCACTATCCCATATCACTGCCAGAGAGATGATGAGATGAGAGCTTGATCATGT 4005
 Db 612 AGACCACTATCCCATATCACTGCCAGAGAGATGATGAGATGAGAGCTTGATCATGT 671
 QY 4006 GACTTGACATGATCACTACTGCTCTGAGTGAAGACTCACTGAGAG-CAGTTT 4064
 Db 672 GACTTGACATGATCACTACTGCTCTGAGTGAAGACTCACTGAGAGAGAGAGAGAG 731
 QY 4065 ACCTCAATTTGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4124
 Db 732 ACCTCAATTTGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
 QY 4125 AAGC-AAATTTGAGAACTGACAGGGTGGTGGCTTTTGGGCTTCTAGTCACTC 4182
 Db 792 GAGCAAAATTAACAAAACACAGAGGATGGTGGCTTTTGGGCTTCTAGTCACTC 851
 QY 4183 AGACTGAC-CGCGCCCGCTTCAACAGGGCGCTTTCATATGTTTAAATTTTAA 4241
 Db 852 AAATTTGACTTTCCCAACCTTTCACAGGGCTTTCATATGTTTAAATTTTAA 911
 QY 4242 TGATGATTTAGCTTTTAAATTTCAATCACTACTTCTTGGCTATTTGGTT 4301
 Db 912 TATATATTTAGCTTTTAAATTTCAATCACTACTTCTTGGCTATTTGGTT 971
 QY 4302 TACAAAAACACCACTATCAAGAGTCCGTGCGGACGATTAATGCTGTCGGG 4361
 Db 972 TGCAAAAAAGCCACTATCAAGAGATG-----CTGATGTCGATTAATAATTTGCCAA 1026
 QY 4362 CGTACGCTAACTGAGAGCTTGTGTAACCTTGTGTCAGTGTGCCAACACAT 4421
 Db 1027 ATGTCCATTAATCTGAGATGTAATTTTTCATTTTCCAGTGTGCCAACATTAAT 1086
 QY 4422 TGATGATTTGGGGCTGTGCTCC-----TGCGGTAGACAGAGAGATGGGTACTCTGT 4477
 Db 1087 TGATGATTTGGGGCTGTGCTCCCTTCAACATAGAGTGCAGAGAGATGATGATCTCTGT 1146

QY 4478 TTTGAAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4537
 Db 1147 TTTGAAAGCTGATA-GAATGAGCCCAATTAAGGAGAGGTTGTGCTGTTGTGTGT 1205
 QY 4538 ATCAAGGTATCC-TGTGGCGATGTACATCTGTAGACGGAAGATGACT-CTTCCCTCAT 4595
 Db 1206 ATCAAGGTATCTTGTGAGATGATTAATCACTCTGTAGATTAAGAAATTTGTTCTTCCA 1265
 QY 4596 GGTAAACCCACACCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4653
 Db 1266 TGCAAGCTATTAATCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1325
 QY 4654 AGTGAAGCTTGTAGCCACTGAGAGAC-CCATGTTTCCGTTGTGCTGATGATGATGAT 4712
 Db 1326 AGTGAAGCTTGTAGCCACTGAGAGACCTGTTGTTTGTGCTGATGATGATGATGATGAT 1385
 QY 4713 AGTTGTGCGGTGTTGTTTATTTTCAATGTCATGTCATGTCATGTCATGTCATGTCATG 4772
 Db 1386 GATTGAGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1445
 QY 4773 TAAAGCGCAACCTGCCAGCGGTTAAGATCATGATGATGATGATGATGATGATGATGAT 4827
 Db 1446 GTAAGCAAC-ACCACTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1504
 QY 4828 TGTATGCTGAAG-TTATTTCCAGTTGCTTTTATGAA-----TATCTTGAACAGTAT 4881
 Db 1505 TGTATCAAGAAATTTTATTTCCAGTTACTTTCATGATGATGATGATGATGATGATGAT 1564
 QY 4882 CTTCTTGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4941
 Db 1565 TTTCTTGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1623
 QY 4942 TTTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5001
 Db 1624 TTTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1682
 QY 5002 TGTGTATTAAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5058
 Db 1683 TGTGTG-TAAGATATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1741
 QY 5059 CGGAGAGAGATT 5071
 Db 1742 TCTCAATTAAT 1754

RESULT 15
 AC106760/c 78190 bp DNA linear PRI 27-MAR-2002
 LOCUS
 DEFINITION Homo sapiens chromosome 5 clone RP11-215G15, complete sequence.
 AC106760
 VERSION
 AC106760.2 GI:19747162
 KEYWORDS
 HTG.
 SOURCE
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 78190)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE
 Direct Submission
 JOURNAL
 Unpublished
 2 (bases 1 to 78190)
 REFERENCE
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (12-JUN-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 78190)
 REFERENCE
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (27-MAR-2002) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Mar 27, 2002 this sequence version replaced gi:18139310.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.7% of Sequence;
 Estimated Total Number of Errors is 0.3.
 NOTE: This insert is not the entire sequence of the clone (entire
 sequence is 179,4kb). It is clipped at the overlap with AC012629.
 The number of bases overlapped is 5036.

FEATURES

source

1. 78190

/organism="Homo sapiens"

/db xref="taxon:9606"

/chromosome="5"

/clone="RP11-215G15"

BASE COUNT 19148 a 18124 c 19405 g 21513 t

ORIGIN

Query Match

Best Local Similarity 88.6%; Score 813.6; DB 9; Length 78190;

Matches 927; Conservative 0; Mismatches 114; Indels 5; Gaps 4;

QY 3225 GGGCCGCAAGCCAGCCATTCGCGAGAGAGTTTATTCGATGACATACGAGAGTTC 3284
 Db 73571 GACTGCTAACCCAGTCGTTTCGGAGAGATTATTCGATGACATACGAGAGTTC 73512
 QY 3285 TGAGCAAGAGATTAACTTTACGACAGGGGATGTGATTGTTACCAAGAAAGATG 3344
 Db 73511 TGAGCAAGAGATTAACTTTACGACAGAGAGATGTGATTGTTACCAAG-AGATGG 73453
 QY 3345 TGACTGGTGACGGGAAAGGTTGGGACCAAGTCCGAGATCTTCCCTTCTAATATGTGAG 3404
 Db 73452 TGACCGGTGACAGAGAAAGATGGGCGACAGCGGATCTTCCCTTCTAATATGTGAG 73393
 QY 3405 GCTTAAAGATTACAGAGGCTCTGAAGTCT-GGGAAACAGGAGTTTAAAGAAAGAAC 3463
 Db 73392 GCTTAAAGATTACAGAGGCTCTGAAGTCTGGGAAATGAGGAGTTTAAAGAAAGAAC 73333
 QY 3464 CTGAATTTGCCAGATTATTCCTACGCTGCTACTGGTCCGAAACAATCACCCTGG 3523
 Db 73332 CTGAATTTGCCAGATTATTCCTACATACACGCGCAGCGGCGGAGCAGCTCCTCTCG 73273
 QY 3524 CTCTGGGCGAGCTGATTTCTATCCGGAAGAAAGCCAGGTGATGTGGAAAGAGAAC 3583
 Db 73272 CCCTGGTCACTGATTTTATCTGAAAAAGAACCAAGGTGATGTGGAAAGAGAAC 73213
 QY 3584 TGCAGCTCAGGAGAAAAAGCCAGATAGGTTGTTCCAGCAATTAATGTCAAACTTC 3643
 Db 73212 TGCAGCTCAGGAGAAAAAGCCAGATAGGTTGTTCCAGCAATTAATGTCAAACTTC 73153
 QY 3644 TAAAGCCCGGAGCAAGCAAAATCAACCCCACTAGCTTACCAAGCCGACAGCCAG 3703
 Db 73152 TAAAGCCCGGAGCAAGCAAAATCACTCCACAGCCACTAGCTTACCAAGCCAG 73093
 QY 3704 CAGTGTCCAGGTGATGCGGATGATGATTAACCCGCGGAGAGATGAGAGTACCTG 3763
 Db 73092 CAGTGTCCAGGTGATGCGGATGATGATTAACCCGCGGAGAGATGAGAGTACCTG 73033
 QY 3764 TCAGCAAGGCGCAGATCATCAAGCTCTCAACAAGAGAGACCCGGACTGGTGGAAAGAG 3823
 Db 73032 TCAGCAAGGCGCAGATCATCAAGCTCTCAACAAGAGAGACCCGGACTGGTGGAAAGAG 72973
 QY 3824 AAGTCACTGGGCAAGTTGGGCTCTCCATCCAAATTAATGTAAGCTGACCAAGACATGG 3883
 Db 72972 AAGTCACTGGGCAAGTTGGGCTCTCCATCCAAATTAATGTAAGCTGACCAAGACATGG 72913
 QY 3884 ACCCGAGCGAGCAATGATATATGTTGTCATCCCGCCCTCAAGGCTGAAAGTCTCTCAA 3943
 Db 72912 ACCCGAGCGAGCAATGATATATGTTGTCATCCCGCCCTCAAGGCTGAAAGTCTCTCAA 72853
 QY 3944 AGAGACCACTATCCATATCACTGCCAGAGGAGATGAGGAGATGACCTTGTATCAT 4003
 Db 72852 AGAGACCACTATCCATATCACTGCCAGAGGAGATGAGGAGATGACCTTGTATCAT 72793
 QY 4004 GTGACTTGACAGATGATCAGCTTCTGAGTGAAGAACTCACTGACAGCAGTT 4063

Db 72792 GTGACTTGACAGATGATCAGCTTCTGAGTGAAGAACTCACTGACAGCAGTT 72733
 QY 4064 TACCTCATTTGACCTTATGATGATGATGAAATGTCTGAGTCACTGCGTGAAGGCA 4123
 Db 72732 TACCTCATTTGACCTTATGATGATGATGAAATGTCTGAGTCACTGCGTGAAGGCA 72673
 QY 4124 GAAGC-AAATGTGAGAACTGCACAGGAGTGGTCCCTTTGGGCTTCTTACGACCT 4181
 Db 72672 GAGCCAAAATTAACAAAATACACAGGAGAGGAGTCTTTGTTGCTTCTTACGACCT 72613
 QY 4182 CAGACTGAC-CGAGCCCGCCTTCAACAGGCGCTTTCAATAGTTTAAAGATTATTTTAA 4240
 Db 72612 CAATTTGACTTTCCCGCAGCTTTGACAGGAGTGTTCATATAGTTTAAAGTTATTTTAA 72553
 QY 4241 ATGTGATTTTACCTTTTAAATAAA 4266
 Db 72552 ATATATTTTAACTTTTAAATAAA 72527

Search completed: March 14, 2003, 08:16:31
 Job time : 12832 secs

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 XX
 PS Claim 6; Page 38-40; 99pp; English.
 XX
 CC The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC
 SQ Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 other;
 XX
 Query Match 99.5%; Score 5060; DB 21; Length 5082;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5082; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Db 541 CCTCCACACTTCCCTGTCATGAAACAGCAACCAAGTGGCTATTCCAGTGCACGCA 600
 Qy TTTGGTATAGAGGATGTTGCTAGCATGCGACCACTCCACAGCTGTGCTGTCATG 660
 Db 601 TTTGGTATAGAGGATGTTGCTAGCATGCGACCACTCCACAGCTGTGCTGTCATG 660
 Qy 661 GGCTCCATTCCAGTTGTTGGAATGTTCCACCTTATGATTTCTGTCCCTCCAGCA 720
 Db 661 GGCTCCATTCCAGTTGTTGGAATGTTCCACCTTATGATTTCTGTCCCTCCAGCA 720
 Qy 721 GTCCTCCCTGGCTAAAGGGGGCTCCCTCCGTCATACAGCTGCTGGCTGGCAT 780
 Db 721 GTCCTCCCTGGCTAAAGGGGGCTCCCTCCGTCATACAGCTGCTGGCTGGCAT 780
 Qy 781 CTTGAGCCACATGGCCAAAGAGTTCTTCTTCAGCAAGATCTGCTCCAGGCTCAAT 840
 Db 781 CTTGAGCCACATGGCCAAAGAGTTCTTCTTCAGCAAGATCTGCTCCAGGCTCAAT 840
 Qy 841 AACATAGTTACAGAAAGGACATTCATGATGTCGCGAGCGCCCTCCAGCAGCAGAA 900
 Db 841 AACATAGTTACAGAAAGGACATTCATGATGTCGCGAGCGCCCTCCAGCAGCAGAA 900
 Qy 901 TGGCTGTGCTCAGTCATCAAGGCTGAAATACAGGAGTTATTCAACAGCCACGACAA 960
 Db 901 TGGCTGTGCTCAGTCATCAAGGCTGAAATACAGGAGTTATTCAACAGCCACGACAA 960
 Qy 961 ACTATGATGACACTTAAACAGGTCCTCCAGCAAGAACTATTCTCATGCAATCAAGTT 1020
 Db 961 ACTATGATGACACTTAAACAGGTCCTCCAGCAAGAACTATTCTCATGCAATCAAGTT 1020
 Qy 1021 CCCAGGCTCAGCTGGCTTCAATATGAAATCTTCTGACATTGATCAAGATGAAAACTC 1080
 Db 1021 CCCAGGCTCAGCTGGCTTCAATATGAAATCTTCTGACATTGATCAAGATGAAAACTC 1080
 Qy 1081 ACTGCAAGAAATTTATCTAGCTATGACCTAATGATGTTGTCATGTCGTGACCCA 1140
 Db 1081 ACTGCAAGAAATTTATCTAGCTATGACCTAATGATGTTGTCATGTCGTGACCCA 1140
 Qy 1141 CTGCGCGCGCTGCTGCTCCGCAATATCATCCCTCTTCTTCAAGAAAGTTGCTCCGGC 1200
 Db 1141 CTGCGCGCGCTGCTGCTCCGCAATATCATCCCTCTTCTTCAAGAAAGTTGCTCCGGC 1200
 Qy 1201 AGTGGATGTCCTGATTAAGCTCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGTCG 1260
 Db 1201 AGTGGATGTCCTGATTAAGCTCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGTCG 1260
 Qy 1261 TCAGAGATAGACAGCCAGCAAGAAAGAACTGCTGTGACATTTGAAGATTAAGAACCG 1320
 Db 1261 TCAGAGATAGACAGCCAGCAAGAAAGAACTGCTGTGACATTTGAAGATTAAGAACCG 1320
 Qy 1321 GAGAACTTCGAGCAGAGGAGTGTGAGCTGAGAGAGCGCCCAAGCGCTCTTGAGCAG 1380
 Db 1321 GAGAACTTCGAGCAGAGGAGTGTGAGCTGAGAGAGCGCCCAAGCGCTCTTGAGCAG 1380
 Qy 1381 CAGCGCAAGAGCGAGAGCGTGTGCTCAGCTGAGAGCGCCCAAGCGCTCTTGAGCAG 1440
 Db 1381 CAGCGCAAGAGCGAGAGCGTGTGCTCAGCTGAGAGCGCCCAAGCGCTCTTGAGCAG 1440
 Qy 1441 CGGAGAGCGCCAGAGAGCGAGAGCGGAGCTGAGCTGAGAGAGCGCTGAGAGAG 1500
 Db 1441 CGGAGAGCGCCAGAGAGCGAGAGCGGAGCTGAGCTGAGAGAGCGCTGAGAGAG 1500
 Qy 1501 CAGCGGAGCTGAGCGGCGCAGAGAGAGAGAGAGAGAGATGAGAGCGCGGAG 1560
 Db 1501 CAGCGGAGCTGAGCGGCGCAGAGAGAGAGAGAGAGAGATGAGAGCGCGGAG 1560
 Qy 1561 GCGCGAAGACGGGAACTGGAAAGCGCAGCACTTGAATGGGAACGGAGACGGAGCAG 1620
 Db 1561 GCGCGAAGACGGGAACTGGAAAGCGCAGCACTTGAATGGGAACGGAGACGGAGCAG 1620
 Qy 1621 GAACTCTGAAATCAGAGAAACAAGAGACAGAGGAGCCTGTGCTCTGAAGGCAAGAGG 1680
 Db 1621 GAACTCTGAAATCAGAGAAACAAGAGACAGAGGAGCCTGTGCTCTGAAGGCAAGAGG 1680

QY 1681 AAGACTCTGGAGTTGATGATTAGAAAGCTCTGAATGACAAAAGCATGCTAGAAAGGAAA 1740
DB 1681 AAGACTCTGGAGTTGATGATTAGAAAGCTCTGAATGACAAAAGCATGCTAGAAAGGAAA 1740
QY 1741 CTTTCAGGATATCAGGTGTGCACTGCGCAACCCAGAGGCAAGAAATTGAGAGCAAGCAAG 1800
DB 1741 CTTTCAGGATATCAGGTGTGCACTGCGCAACCCAGAGGCAAGAAATTGAGAGCAAGCAAG 1800
QY 1801 TCTAGAGAGTAAAGATGTGTAATCACCACCTTACAGAGAGTGTGAGGAATCTGAG 1860
DB 1801 TCTAGAGAGTAAAGATGTGTAATCACCACCTTACAGAGAGTGTGAGGAATCTGAG 1860
QY 1861 CAATCTCTGGAGAGCTTATTCAGAGAAAACAGATATCTAGTACAGGATTAACAAAGTC 1920
DB 1861 CAATCTCTGGAGAGCTTATTCAGAGAAAACAGATATCTAGTACAGGATTAACAAAGTC 1920
QY 1921 CAGCAGAAACGTTTGCATAGAGACTGCTTCTTACCTTCAAAAAGAGCTTTGAGAACAG 1980
DB 1921 CAGCAGAAACGTTTGCATAGAGACTGCTTCTTACCTTCAAAAAGAGCTTTGAGAACAG 1980
QY 1981 GAGCTGCCCCGAGAGAGCTCCGAGAGAGCTGAGAGTGTGAGAGAGAGAGAGAGAGTCA 2040
DB 1981 GAGCTGCCCCGAGAGAGCTCCGAGAGAGCTGAGAGTGTGAGAGAGAGAGAGAGAGTCA 2040
QY 2041 AAGCTGAGAGAGATGATGTTTCAACAAACAGCTGAGAGAACTGAGAGAGATACATAGC 2100
DB 2041 AAGCTGAGAGAGATGATGTTTCAACAAACAGCTGAGAGAACTGAGAGAGATACATAGC 2100
QY 2101 AAAACAGAACTCCAGAAAGCAGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB 2101 AAAACAGAACTCCAGAAAGCAGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 GAGAGAGAGAGCTGAGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
DB 2161 GAGAGAGAGAGCTGAGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2221 AAGGAGCAAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
DB 2221 AAGGAGCAAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 CACGAG 2340
DB 2281 CACGAG 2340
QY 2341 GCCAAGCCGGAATGCAAG 2400
DB 2341 GCCAAGCCGGAATGCAAG 2400
QY 2401 AAGCTGAG 2460
DB 2401 AAGCTGAG 2460
QY 2461 CAGAGAGAGTAAAGTGTATATATACCGAGCGCTGTATCCCTTTGATCCAGAGTCAAC 2520
DB 2461 CAGAGAGAGTAAAGTGTATATATACCGAGCGCTGTATCCCTTTGATCCAGAGTCAAC 2520
QY 2521 GATGAGATTCACATCCAG 2580
DB 2521 GATGAGATTCACATCCAG 2580
QY 2581 CCAGAGATGCTGAG 2640
DB 2581 CCAGAGATGCTGAG 2640
QY 2641 GAAAAAGATTCAGAAAAATGAGGTTCCATCTCAGCCCAACAGAGTCAATCTGACATCT 2700
DB 2641 GAAAAAGATTCAGAAAAATGAGGTTCCATCTCAGCCCAACAGAGTCAATCTGACATCT 2700
QY 2701 GCGCTGCCCCCAAACTGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
DB 2701 GCGCTGCCCCCAAACTGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760

QY 2761 GAGCCCTTCACACACCCCAACCACTGGGAGAGCTTCACTTCCAGCTGAGCCAGAGAGTCA 2820
DB 2761 GAGCCCTTCACACACCCCAACCACTGGGAGAGCTTCACTTCCAGCTGAGCCAGAGAGTCA 2820
QY 2821 AAG 2880
DB 2821 AAG 2880
QY 2881 CCTAGTGTGAG 2940
DB 2881 CCTAGTGTGAG 2940
QY 2941 TCCCATCTCCCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
DB 2941 TCCCATCTCCCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
QY 3001 TATCCCTGAG 3060
DB 3001 TATCCCTGAG 3060
QY 3061 GTTCTGAG 3120
DB 3061 GTTCTGAG 3120
QY 3121 CCAAGCTTCTAG 3180
DB 3121 CCAAGCTTCTAG 3180
QY 3181 GAGCCCTTCTAG 3240
DB 3181 GAGCCCTTCTAG 3240
QY 3241 ATTCCGAG 3300
DB 3241 ATTCCGAG 3300
QY 3301 ACCCTTCAAG 3360
DB 3301 ACCCTTCAAG 3360
QY 3361 ACGGTGAG 3420
DB 3361 ACGGTGAG 3420
QY 3421 GAGCTGAG 3480
DB 3421 GAGCTGAG 3480
QY 3481 ATTGCTTCTAG 3540
DB 3481 ATTGCTTCTAG 3540
QY 3541 CTGATCCGAG 3600
DB 3541 CTGATCCGAG 3600
QY 3601 AAGCCGAG 3660
DB 3601 AAGCCGAG 3660
QY 3661 AAAATCAG 3720
DB 3661 AAAATCAG 3720
QY 3721 GGGATGAG 3780
DB 3721 GGGATGAG 3780
QY 3781 ATCAAGCTCTCAAG 3840
DB 3781 ATCAAGCTCTCAAG 3840
QY 3841 GGGCTTCTCCATCAATTAATTAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
DB 3841 GGGCTTCTCCATCAATTAATTAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900

```

Db 3840 GGGGCTTCCCATCCAAATTAATGTAAGCTGACCAAGACATGAGCCAGCCAGCCAAATGA 3899
Qy 3901 ATCAATATGTGTCATCCCGCCCTGAGGCTGGAAGTCTCTAAAGAGACCACATATCCA 3960
Db 3900 ATCAATATGTGTCATCCCGCCCTGAGGCTGGAAGTCTCTAAAGAGACCACATATCCA 3959
Qy 3961 TATCACTGCCAGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATG 4020
Db 3960 TATCACTGCCAGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 4019
Qy 4021 CACCTACGCTCTTCTGATGATAAAGACCTGACGAGAGACGAGATTAATCTTACCTTA 4080
Db 4020 CACCTACGCTCTTCTGATGATAAAGACCTGACGAGAGACGAGATTAATCTTACCTTA 4079
Qy 4081 GTTGATGTGATCGAAATGTCAGTCACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 4080 GTTGATGTGATCGAAATGTCAGTCACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 4139
Qy 4141 TGCAAGGGGTGGGTGCTTTTGGGGCTTCTGATGATCACTGAGACCGGCGCCGCC 4200
Db 4140 TGCAAGGGGTGGGTGCTTTTGGGGCTTCTGATGATCACTGAGACCGGCGCCGCC 4199
Qy 4201 TTCAACAGGGGCTTCAATAGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4260
Db 4200 TTCAACAGGGGCTTCAATAGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4259
Qy 4261 ATAAATCTCAATCAATCAATCTTCTGCTATTTTGTGTTTAAATTTTAAATTTTAAAT 4320
Db 4260 ATAAATCTCAATCAATCAATCTTCTGCTATTTTGTGTTTAAATTTTAAATTTTAAATTT 4319
Qy 4321 AAGAGAGGCTGCTGCGGAGAGATTTAAATGCTGTTCGGGGCTGACCGTAACTGAGAGC 4380
Db 4320 AAGAGAGGCTGCTGCGGAGAGATTTAAATGCTGTTCGGGGCTGACCGTAACTGAGAGC 4379
Qy 4381 TTGCTGATCCTTTGCGGCTTGTCCAGTGTTCACCAACCAATTTGATGAGAGCTTT 4440
Db 4380 TTGCTGATCCTTTGCGGCTTGTCCAGTGTTCACCAACCAATTTGATGAGAGCTTT 4439
Qy 4441 CCTGCGGTAGAGACAGAGAGATGGGTGATCCTGTTTGAATTTGATGATGATGATG 4500
Db 4440 CCTGCGGTAGAGACAGAGAGATGGGTGATCCTGTTTGAATTTGATGATGATGATG 4499
Qy 4501 AGCCTGACTATGAAAGGGGTATGCTGTGATGACCATGACGATGACCTGTGGCGCATG 4560
Db 4500 AGCCTGACTATGAAAGGGGTATGCTGTGATGACCATGACGATGACCTGTGGCGCATG 4559
Qy 4561 TACCATCTGTAACGAGAGATGCTTCTCTCATGCTAAACCCACACCGGTGATCAGT 4620
Db 4560 TACCATCTGTAACGAGAGATGCTTCTCTCATGCTAAACCCACACCGGTGATCAGT 4619
Qy 4621 GCTCTCATCTACCTGATTCATTTTATTTGACAGTACCTTTGAGCAGCAGCAGCAGCAG 4680
Db 4620 GCTCTCATCTACCTGATTCATTTTATTTGAGCAGTACCTTTGAGCAGCAGCAGCAGCAG 4678
Qy 4681 ACCCATGTTCCGTTGGTGTCAATGATGATGATGATGATGATGATGATGATGATGAT 4740
Db 4679 ACCCATGTTCCGTTGGTGTCAATGATGATGATGATGATGATGATGATGATGATGAT 4738
Qy 4741 TCAATCTGGCATGTTTCAACCATTAACCTAGTAGACGCCCACTGCCAGCGGTTACG 4800
Db 4739 TCAATCTGGCATGTTTCAACCATTAACCTAGTAGACGCCCACTGCCAGCGGTTACG 4798
Qy 4801 ATCATGATACCCACCGCTTAGCTGCTGTAAGTAAAGTTTATTCAGTGTGCTTTTAT 4860
Db 4799 ATCATGATACCCACCGCTTAGCTGCTGTAAGTAAAGTTTATTCAGTGTGCTTTTAT 4858
Qy 4861 GGAATATCTTGAACAGTAATCTTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920
Db 4859 GGAATATCTTGAACAGTAATCTTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4918
Qy 4921 TTAATTTCCAGGTTTACATTTTAACTAGACTGTGGGGGTTGCTACGATTAATATG 4980

```

```

Db 4919 TTAATTTCCAGGTTTACATTTTAACTAGACTGTGGGGGTTGCTACAGATTAATATG 4978
Qy 4981 AATGGCGCTCTGAGCGCTGTGTTAACTTGTGCTGATGAGAGAGAGAGAGAGAGAGAG 5040
Db 4979 AATGGCGCTCTGAGCGCTGTGTTAACTTGTGCTGATGAGAGAGAGAGAGAGAGAG 5038
Qy 5041 TAGATATTAGTTGGAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5084
Db 5039 TAGATATTAGTTGGAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5082

RESULT 2
ID AA239024 standard; cDNA; 5738 BP.
XX
AC AA239024;
XX
DT 28-FEB-2000 (first entry)
XX
DE Mouse Esei1 cDNA sequence.
XX
KW Mouse; murine; Esei1; Esei2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
OS Mus sp.
XX
PN W09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
XX
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Egan SE, Wang W, Sengar A;
XX
DR WPI, 2000-052802/04.
XX
DR P-Psdb; AAY57449.
XX
PT New nucleic acid encoding Esei1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection.
XX
PS Claim 6; Page 56-59; 99p; English.
XX
CC The present invention specifically describes mammalian Esei1 and 2
CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-
CC mediated endocytosis (as a complex with Eps15 protein), vesicular
CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC sequences antisense to the (I) polynucleotide; agents that downregulate
CC expression of Ese genes or antagonists of an Ese binding partner are
CC used to treat diseases associated with unstable endocytosis and
CC resulting changes in cellular function. Particularly overexpression of
CC Esei1 is used to block clathrin-mediated endocytosis in vivo or in cell
CC cultures, while administration of (I) is used to promote endocytosis of
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC proliferation of cells that can be stimulated to proliferate by a growth
CC factor receptor; and similar compounds (also inactive Ese mutants) can be
CC used to prevent viral infection. Endocytosis may also be regulated, in
CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
CC dynamin to the complex. Generally conditions that can be treated include
CC cancer; abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission. The
CC present sequence represents mouse Esei1 cDNA sequence.
XX
SQ Sequence 5738 BP; 1556 A; 1456 C; 1613 G; 1113 T; 0 other;

```


PA (HSCR-) HSC RES & DEV LP.
 XX Egan SE, Wang W, Sengar A;
 XX MPI: 2000-052802/04.
 DR P-PSDB; AAY57444.
 XX
 XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 XX Claim 6: Page 40-42; 9pp; English.
 XX
 XX The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ise = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (anti)agonists, mimetics, fragments and inactive
 CC mutants); (I) -specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Anti)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX
 XX Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;
 SQ
 Query Match 73.2%; Score 3723; DB 21; Length 3723;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGGGCTCTCCCTCATACAGAGCTCTGCTGCTTTGGCATCTCTGACGACATGGCCA 798
 Db 481 GGGGCTCTCCCTCATACAGAGCTCTGCTGCTTTGGCATCTCTGACGACATGGCCA 540
 QY 799 AAGAGTCTCTCCCTCAGCAGATCGTGGTCCAGGGTCAAAATTAACAATAAGTTACAGAG 858
 Db 541 AAGAGTCTCTCCCTCAGCAGATCGTGGTCCAGGGTCAAAATTAACAATAAGTTACAGAG 600
 QY 859 GCACAAATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAGAAATGGGCTGCTCACTCA 918
 Db 601 GCACAAATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAGAAATGGGCTGCTCACTCA 660
 QY 919 TCAAGCTGAATACAGCAGATTAATTAACAGCCACGACAAACTATAGATGACACTTA 978
 Db 661 TCAAGCTGAATACAGCAGATTAATTAACAGCCACGACAAACTATAGATGACACTTA 720
 QY 979 ACAGTCCCCGAGGAAGAACTATCTCATGCAATCAAGTTTACCCGAGGCTCAGCTGGCT 1038
 Db 721 ACAGTCCCCGAGGAAGAACTATCTCATGCAATCAAGTTTACCCGAGGCTCAGCTGGCT 780
 QY 1039 TCAATATGAAATCTTTCTGACATTTGATCAAGATGAAAACTCACTGACAGAAATTTATC 1098
 Db 781 TCAATATGAAATCTTTCTGACATTTGATCAAGATGAAAACTCACTGACAGAAATTTATC 840
 QY 1099 CTAGCTATGACCTTAATGATGTTGATGTCATGTCAGCCACTGCGCCGCTCTGCT 1158
 Db 841 CTAGCTATGACCTTAATGATGTTGATGTCATGTCAGCCACTGCGCCGCTCTGCTCT 900
 QY 1159 CCAGAAATCATCT 1218
 Db 901 CCAGAAATCATCT 960
 QY 1219 AGCTCTTTCTTGATGATCAGAGGCTCTGAGGAGCGCTGTCAGAGATGAGCAGCAG 1278
 Db 961 AGCTCTTTCTTGATGATCAGAGGCTCTGAGGAGCGCTGTCAGAGATGAGCAGCAG 1020
 QY 1279 CCAGAGAAAGAACTGCTCTGATCACTTTGAAGATTAAGAGCGGGAGAACTTGCAGCGC 1338
 Db 1021 CCAGAGAAAGAACTGCTCTGATCACTTTGAAGATTAAGAGCGGGAGAACTTGCAGCGC 1080
 QY 1339 AGTGTGAGCTGAGAAAGCGCGCCAGCGCTCTTGAAGCAGCGCCAAAGACAGCAGAG 1398
 Db 1081 AGTGTGAGCTGAGAAAGCGCGCCAGCGCTCTTGAAGCAGCGCCAAAGACAGCAGAG 1140
 QY 1399 CGGTGGCTCAGCTGAGAGCGCGCCAGCAGAGAGAGAAAGCGGGAGCCGACAGAGCAG 1458
 Db 1141 CGGTGGCTCAGCTGAGAGCGCGCCAGCAGAGAGAGAAAGCGGGAGCCGACAGAGCAG 1200
 QY 1459 GAGGCGAAGCGGAGCTGAGCTGAGAGAGCGCTGAGAGAGCGGGAGCTGAGAGCGG 1518
 Db 1201 GAGGCGAAGCGGAGCTGAGCTGAGAGAGCGCTGAGAGAGCGGGAGCTGAGAGCGG 1260
 QY 1519 CAGCGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCCGACAAACGGGAACTG 1578
 Db 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCCGACAAACGGGAACTG 1320
 QY 1579 GAAAGGAGCGAGCAACTTGAATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
 Db 1321 GAAAGGAGCGAGCAACTTGAATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1639 AACAG 1698
 Db 1381 AACAG 1440
 QY 1699 TTAAGAGCTCTGAGATGACAAAGAGATGAGTAAAGAGAACTTCAAGATAGAGGT 1758
 Db 1441 TTAAGAGCTCTGAGATGACAAAGAGATGAGTAAAGAGAACTTCAAGATAGAGGT 1500
 QY 1759 CGACTGGCAACCCAGAGGCAAGAAATGAGAGCAGAACTAGAGAGCTAAGAAAT 1818
 Db 1501 CGACTGGCAACCCAGAGGCAAGAAATGAGAGCAGAACTAGAGAGCTAAGAAAT 1560
 QY 1819 GCTGAATACACCACCTTACAGCAGAGAGTTGAGAGAAATCTAGCAAAATGCTTGAAGACT 1878

Db 1561 GCTGAATCAACCACTTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGGAAACCTT 1620
Qy 1879 ATTCCAGAGAAAACAGATACCTCAGTGAACCACTTAAACCAAGTCCAGCAAGAACAGTTGGCAT 1938
Db 1621 ATTCCAGAGAAAACAGATACCTCAGTGAACCAAGTCCAGCAAGAACAGTTGGCAT 1680
Qy 1939 AGAGACTCGCTTCTTACCTCCAAAGAGCCTTGGAGCAAGAGAGCTGGCCCGCAGCAG 1998
Db 1681 AGAGACTCGCTTCTTACCTCCAAAGAGCCTTGGAGCAAGAGAGCTGGCCCGCAGCAG 1740
Qy 1999 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACAGCTCAAGCTCGACGAGATTGAT 2058
Db 1741 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACAGCTCAAGCTCGACGAGATTGAT 1800
Qy 2059 GTTTTCAACCAACAGCTGGAAGAACTGAGAGATATCATGCAAAACAGCAACTCCGAAG 2118
Db 1801 GTTTTCAACCAACAGCTGGAAGAACTGAGAGATATCATGCAAAACAGCAACTCCGAAG 1860
Qy 2119 CAGAGTCCCTGAGCAGCAGCGCAGCTGAGCAGAAAGAGAGAGAGAGAGCCTGGAG 2178
Db 1861 CAGAGTCCCTGAGCAGCAGCGCAGCTGAGCAGAAAGAGAGAGAGAGAGCCTGGAG 1920
Qy 2179 TTAG 2238
Db 1921 TTAG 1980
Qy 2239 GAGCATGTGAGCAG 2298
Db 1981 GAGCATGTGAGCAG 2040
Qy 2299 AAGAGGAG 2358
Db 2041 AAGAGGAG 2100
Qy 2359 GACAG 2418
Db 2101 GACAG 2160
Qy 2419 CCTGTGCTACACAG 2478
Db 2161 CCTGTGCTACACAG 2220
Qy 2479 GTATATTACCGAG 2538
Db 2221 GTATATTACCGAG 2280
Qy 2539 CAGAGAGATTAAGTCAATGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2598
Db 2281 CAGAGAGATTAAGTCAATGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy 2599 GAGCTGAAAG 2658
Db 2341 GAGCTGAAAG 2400
Qy 2659 GAGCTTCCCACTCCAGCAGCAACAGTACAGTACATCTGAGCCTTGGCCCAAACTG 2718
Db 2401 GAGCTTCCCACTCCAGCAGCAACAGTACAGTACATCTGAGCCTTGGCCCAAACTG 2460
Qy 2719 GCTTGTGCTGAGAGCCTGTCTCTTTGCAAGTGAAGCTTTTCTGAGCCTTCCCAACCTCC 2778
Db 2461 GCTTGTGCTGAGAGCCTGTCTCTTTGCAAGTGAAGCTTTTCTGAGCCTTCCCAACCTCC 2520
Qy 2779 AACAACTGGGAG 2838
Db 2521 AACAACTGGGAG 2580
Qy 2839 GACAACTGGGAG 2898
Db 2581 GACAACTGGGAG 2640
Qy 2899 CGGAGAGATCAGCCTTACCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2958

Db 2641 CGGAGAGATCAGCCTTACCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2700
Qy 2959 GGGCAGGGGTGAAAAGGTGAAAGGGCTACAAGCCGCAACCCCTGTATCCCTGAGAGCCAAA 3018
Db 2701 GGGCAGGGGTGAAAAGGTGAAAGGGCTACAAGCCGCAACCCCTGTATCCCTGAGAGCCAAA 2760
Qy 3019 AAAGCAACCACTTAAATTTTAAACAAAAGTGAAGCTATCAACCTTCTGAAACAGCAAGAC 3078
Db 2761 AAAGCAACCACTTAAATTTTAAACAAAAGTGAAGCTATCAACCTTCTGAAACAGCAAGAC 2820
Qy 3079 ATGTGAGTGTGAG 3138
Db 2821 ATGTGAGTGTGAG 2880
Qy 3139 CTCAATTTGAGGGCCGTAAAGGAAATCCACAGCATGATCTGAGCCTTCTGAAAGTCT 3198
Db 2881 CTCAATTTGAGGGCCGTAAAGGAAATCCACAGCATGATCTGAGCCTTCTGAAAGTCT 2940
Qy 3199 GCTAGCTTAAAG 3258
Db 2941 GCTAGCTTAAAG 3000
Qy 3259 ATTGCATGTACATACAG 3318
Db 3001 ATTGCATGTACATACAG 3060
Qy 3319 GTGATTTGTGTTACCAAGAAAGATGTTGATCTGTGAGCAGGAAAGAGTGGCCGACAGTCC 3378
Db 3061 GTGATTTGTGTTACCAAGAAAGATGTTGATCTGTGAGCAGGAAAGAGTGGCCGACAGTCC 3120
Qy 3379 GAGTCTTCCCTTCTAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3438
Db 3121 GAGTCTTCCCTTCTAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Qy 3439 AAAACAGGAGATTAGAGAAAAAACTGAATTTGCCAGGTTATTTGCTTCTAGAGTCT 3498
Db 3181 AAAACAGGAGATTAGAGAAAAAACTGAATTTGCCAGGTTATTTGCTTCTAGAGTCT 3240
Qy 3499 ACTGTTCCCGAACAACCTACCTTGGCTCTGGGAGAGTGAATCTGATCCGGAAGAAAGAAC 3558
Db 3241 ACTGTTCCCGAACAACCTACCTTGGCTCTGGGAGAGTGAATCTGATCCGGAAGAAAGAAC 3300
Qy 3559 CCAAGTGTGATGTTGAG 3618
Db 3301 CCAAGTGTGATGTTGAG 3360
Qy 3619 TTTCCAGCAAAATTAATGTCAAACTTTAAGCCCGGAAACAAAGCAAAATACCCCAACTGAG 3678
Db 3361 TTTCCAGCAAAATTAATGTCAAACTTTAAGCCCGGAAACAAAGCAAAATACCCCAACTGAG 3420
Qy 3679 CTACCCAGAGACCGAG 3738
Db 3421 CTACCCAGAGACCGAG 3480
Qy 3739 GAGAGCCTGGAGCTGGTGAAG 3798
Db 3481 GAGAGCCTGGAGCTGGTGAAG 3540
Qy 3799 GAGAGCCTGGAGCTGGTGAAG 3858
Db 3541 GAGAGCCTGGAGCTGGTGAAG 3600
Qy 3859 TATGTAAAGCTGACCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3918
Db 3601 TATGTAAAGCTGACCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Qy 3919 CCCCCTCAGGCTTGAAGTCTCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3978
Db 3661 CCCCCTCAGGCTTGAAGTCTCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Qy 3979 TGA 3981
Db 3721 TGA 3723

RESULT 4
AA239025
ID AA239025 standard; cDNA; 5144 BP.
XX
XX AA239025;
AC
XX 28-FEB-2000 (first entry)
XX
XX Mouse Eesll coding sequence.
DE
XX Mouse; murine; Eesl; Ees2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
XX Mus sp.
OS
XX WO9955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
PI
XX MPI; 2000-052802/04.
XX P-PSDB; AAY57449.
XX
XX New nucleic acid encoding Eesl and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
XX
PS Claim 6; Page 59-62; 99pp; English.
XX
XX The present invention specifically describes mammalian Eesl and 2
XX proteins (1) and their splice variants (Ese = EH-domain and SH3-domain
XX regulator of endocytosis). (1) are involved in regulation of clathrin-
XX mediated endocytosis (as a complex with Eespl protein), vesicular
XX trafficking and actin cytoskeleton. Generally (1) (or its (ant)agonists,
XX mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab);
XX sequences antisense to the (1) polynucleotide; agents that downregulate
XX expression of Ees genes or antagonists of an Ees binding partner are
XX used to treat diseases associated with undesirable endocytosis and
XX resulting changes in cellular function. Particularly overexpression of
XX Eesl is used to block clathrin-mediated endocytosis in vivo or in cell
XX cultures, while administration of (1) is used to promote endocytosis of
XX selected cells. (ant)agonists of (1) or Ab are used to suppress abnormal
XX proliferation of cells that can be stimulated to proliferate by a growth
XX factor receptor; and similar compounds (also inactive Ees mutants) can be
XX used to prevent viral infection. Endocytosis may also be regulated, in
XX vivo or in cell cultures, by forming an Ese-Eespl complex, then binding
XX dynamin to the complex. Generally conditions that can be treated include
XX cancer; abnormal cell division or migration; viral infection; or abnormal
XX receptor signalling, tissue development or synaptic transmission. The
XX present sequence represents mouse Eesll coding sequence.
XX
XX Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 other;

Query Match 71.6%; Score 3639.4; DB 21; Length 5144;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 ATGGCTAGTTTCCCAACCTTTCGGTGTAGCTGTGATGCTGGGCAATTAAGTGAGAG 318
DB 1 ATGGCTAGTTTCCCAACCTTTCGGTGTAGCTGTGATGCTGGGCAATTAAGTGAGAG 60

QY 319 GAAAGGCGAAGCATGACCAAGAGTCTTACGCTGAAGCCGATAGCGGATTTATTACT 378
DB 61 GAAAGGCGAAGCATGACCAAGAGTCTTACGCTGAAGCCGATAGCGGATTTATTACT 120
QY 379 GGTATCAAGGAGAGAACTTTTTCATCTGGGTTACTAGCCTGCTTAGACAA 438
DB 121 GGTATCAAGGAGAGAACTTTTTCATCTGGGTTACTAGCCTGCTTAGACAA 180
QY 439 ATATGGGCGTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTCCATA 498
DB 181 ATATGGGCGTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTCCATA 240
QY 499 GCCATGAAGCTTATCAAACTGAAGCTTCAAGATATAGCTCCCTCCACACTTCCCT 558
DB 241 GCCATGAAGCTTATCAAACTGAAGCTTCAAGATATAGCTCCCTCCACACTTCCCT 300
QY 559 GTCAATGAAGCAAGCAACAGTGGCTATTTCAGATGACACCACTTTGGTATAGAGGAT 618
DB 301 GTCAATGAAGCAAGCAACAGTGGCTATTTCAGATGACACCACTTTGGTATAGAGGAT 360
QY 619 GCTAGCATGCGCAGCACTACAGAGTGTCTCTGTGCGCAATGGGCTCATTCAGTTGT 678
DB 361 GCTAGCATGCGCAGCACTACAGAGTGTCTCTGTGCGCAATGGGCTCATTCAGTTGT 420
QY 679 GGAATGTCTCCACCTTAGTATCTTCTGTCTCTCCACAGCAGTCTCCCTGGCTAAC 738
DB 421 GGAATGTCTCCACCTTAGTATCTTCTGTCTCTCCACAGCAGTCTCCCTGGCTAAC 480
QY 739 GGGGCTCTCCCTCATACAGGCTCTGCTGCTGGTGGATCTGAGCAGACATGGCCA 798
DB 481 GGGGCTCTCCCTCATACAGGCTCTGCTGCTGGTGGATCTGAGCAGACATGGCCA 540
QY 799 AAGAGTCTTCTCTCAGCAGATCTGTCTCAGAGGTCACATTAATTAAGTATCAGAG 858
DB 541 AAGAGTCTTCTCTCAGCAGATCTGTCTCAGAGGTCACATTAATTAAGTATCAGAG 600
QY 859 GCACATCATTTGATGTGCGCAGCGCCCTTCACAGCAGCAAGTGGCTGTGCTCACTCA 918
DB 601 GCACATCATTTGATGTGCGCAGCGCCCTTCACAGCAGCAAGTGGCTGTGCTCACTCA 660
QY 919 TCAAGGCTGAATACAGGAGTATTAACAAGCCAGCAAACTATGATGAGCACTTA 978
DB 661 TCAAGGCTGAATACAGGAGTATTAACAAGCCAGCAAACTATGATGAGCACTTA 720
QY 979 ACAGTCCCGCAGGAGAGCAATCTCATGTGCAATCAAGTTTACCCAGGCTCAGCTGCT 1038
DB 721 ACAGTCCCGCAGGAGAGCAATCTCATGTGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
QY 1039 TCAATATGAATCTTCTGTGATTCATGATTCAGATGAAGAACTCACTGACAGAAATTTATC 1098
DB 781 TCAATATGAATCTTCTGTGATTCATGATTCAGATGAAGAACTCACTGACAGAAATTTATC 840
QY 1099 CTAGTATGACATTAATGATGATTCATGATTCATGATTCATGATTCATGATTCATGAT 1158
DB 841 CTAGTATGACATTAATGATGATTCATGATTCATGATTCATGATTCATGATTCATGAT 900
QY 1159 CCAGATATCATTCCT 1218
DB 901 CCAGATATCATTCCT 960
QY 1219 AGCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTGATGAGATGAGCAGCAG 1278
DB 961 AGCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTGATGAGATGAGCAGCAG 1020
QY 1279 CCAAGAGAAAGAACTGCTGTGATCAATTAAGAAAGGAGAACTTTCGAGCGAGGC 1338
DB 1021 CCAAGAGAAAGAACTGCTGTGATCAATTAAGAAAGGAGAACTTTCGAGCGAGGC 1080
QY 1339 AGTGTGAGCTGAGAAAGCGCCGCAAGCGCTTTTGAAGAGCAGCCCAAGAGCAGAG 1398
DB 1081 AGTGTGAGCTGAGAAAGCGCCGCAAGCGCTTTTGAAGAGCAGCCCAAGAGCAGAG 1140
QY 1399 CGGTTGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAAAGCGGAGCGCCAGAGAGCAG 1458

QY 3619 TTTCAGCAAAATTATGTCAACTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3678
 CC |||||
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormalities, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions,
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality; myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 CC XX

QY 3679 CTACCCCAAGCCCGAGTGCAGCCAGCAGTGTGCGGATGTACGATTACACC 3738
 DB 3421 CTACCCCAAGCCCGAGTGCAGCCAGCAGTGTGCGGATGTACGATTACACC 3480
 QY 3739 GCCCAGAACATGACGACCTTACGAAAGGCCAGATCTCAACGCTCAACCAAG 3798
 DB 3481 GCCCAGAACATGACGACCTTACGAAAGGCCAGATCTCAACGCTCAACCAAG 3540
 QY 3799 GAGGACCCGAGCTGTGGAAGAAGAAGTCACTGCGCAAGTTGGCTTCCCATCAAT 3858
 DB 3541 GAGGACCCGAGCTGTGGAAGAAGAAGTCACTGCGCAAGTTGGCTTCCCATCAAT 3600
 QY 3859 TATGTAAAGCTGACACACATGACATGACCCGACGCAATG 3899
 DB 3601 TATGTAAAGCTGACACACATGACATGACCCGACGCAATG 3641

RESULT 5
 AAZ34571
 ID AAZ34571 standard; cDNA; 5458 BP.
 XX
 AC AAZ34571;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A cDNA clone 21.
 XX
 SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 267..3929
 FT /tag= a
 XX
 PN WO9953062-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 16-APR-1999; 99MO-US08371.
 XX
 PR 16-APR-1998; 98US-0082007.
 XX
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 PI Korenberg JR, Chen X;
 XX
 DR WPI, 1999-633829/54.
 DR P-PSDB; AAY32155.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 XX
 PS Claim 2; Fig 8; 99pp; English.
 XX
 CC This is the nucleotide sequence of full-length cDNA (clone 21)
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
 CC that contributes to the development of platelets and the
 CC pathogenesis of leukaemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain

CC (see AAZ34570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormalities, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions,
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality; myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 CC XX

Sequence 5458 BP; 1671 A; 1201 C; 1307 G; 1279 T; 0 other;

Query Match 69.4%; Score 3529.4; DB 20; Length 5458;
 Best Local Similarity 83.7%; Pred. No. 0;
 Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;

QY 7 GAGGAGAGTGGAGCGCGCGGAGAGGCGCGCAGCTTGTGCTCCGTAGTACGGCGCT 66
 DB 25 GAGAGAGTGGAGCGCGCGGAGAGGCGCGTACCTTGTGCTCCGTAGTACGGCGCT 84
 QY 67 CGCAAGGAGCATTCGAGCGGCTCCGCGGAGCGCGGAGGAGCGCGCGCGCG 126
 DB 85 CGCGAGGAAGAAATCCGAGCGGCTCCGCGAGC-----GACAGAGAGCGCGCG 133
 QY 127 GGGATGTGTGCGCGCGCTCGGAGCTCGGCTTCTCGC-GCGGAGTGGCGGCTGACATGA 185
 DB 134 GGGATGTGTGCGCGCGCTCGGAGCTCGGCTTCTCGC-CGCGGAGGCGGCTGACGACATGA 193
 QY 186 TTTGTGTGAGGGGCGCGCGCGAGCACCAGCGAGATGAGCGCTGATCAAGAGTGA 245
 DB 194 TTTGTCTCTGGGGCGCGCGCGAGCACCAGCGAGATGAGCGCTGATCAAGAGTGA 253
 QY 246 AGTAATAGAACATGCTGCTAGTTTCCACACCTTGTGCTGAGTGTGCTGGGC 305
 DB 254 AAGTAACAGAACATGCTGCTAGTTTCCACACCTTGTGCTGAGTGTGCTGGGC 313
 QY 306 CATACTGTGAGGAGGAAGGCGCAAGCATGACAGAGTCTTGTGCTGAGCGGATAGC 365
 DB 314 CATACGTGAGGAGGAAGGCGCAAGCATGACAGAGTCTTGTGCTGAGCGGATAGC 373
 QY 366 GGGATTTATTAAGTGTGATCAAGCGAGGAACCTTTTTCATCTGGGTTAAGCTACGCC 425
 DB 374 TGGATTCATTAAGTGTGATCAAGCGAGGAACCTTTTTCATCTGGGTTAAGCTACGCC 433
 QY 426 TGTCTTAACACAAATATGAGGCGCTAGCGAGCATGATATACAGTGAAGATGATCAAGT 485
 DB 434 TGTCTTAACACAGATATGAGGCGCTAGCGAGCATGATATATGATGAAGATGATCAAGT 493
 QY 486 GGAATTTTTCATAGCCATGAAAGCTTATCAAACTGAAGTCAAGATATCAAGTCTCCCTC 545
 DB 494 GGAATTTTTCATAGCCATGAAAGCTTATCAAACTGAAGTCAAGATATCAAGTCTCCCTC 553
 QY 546 CACACTTCCCCCTGTCATGAAAGACAGCAAGCTGATTTTCCAGTGCAGAGATTTGG 605
 DB 554 TGCACCTTCCCCCTGTCATGAAAGACAGCAAGCTGATTTTCTAGCGCACAGCATTTGG 613
 QY 606 TATGAGAGGATTTGCTAGCATGCAACCACTGACAGCTTGTCTCTGTGCAATGGGCTC 665
 DB 614 TATGAGAGGATTTGCTAGCATGCAACCACTGACAGCTTGTCTCTGTGCAATGGGATC 673
 QY 666 CATTCAGTTTGTGAAGTGTCTCACCTTATGATCTTCTCTCTCCAGACAGAGAGTCC 725
 DB 674 CATTCAGTTTGTGAAGTGTCTCACCTTATGATCTTCTCTCTCCAGACAGAGAGTCC 733

XX Homo sapiens.
 OS Key Location/Qualifiers
 FH CDS 208..3642
 FT /tag= a
 XX MO9953062-A2.
 XX 21-OCT-1999.
 XX 16-APR-1999; 99WO-US08371.
 XX 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 PA Korenberg JR, Chen X;
 PI MPI; 1999-633829/54.
 DR P-PSDB; AAY32154.
 XX Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 PS Claim 2; Fig 5; 99pp; English.
 XX This is the nucleotide sequence of full-length cDNA corresponding
 CC to a novel human SH3 gene, termed the SH3D1A gene, that contributes
 CC to the development of platelets and the pathogenesis of Leukaemias,
 CC both in general and in particular those involving the
 CC megakaryocytic lineage. The SH3D1A gene maps to the small
 CC candidate region for low platelets on chromosome 21. Sequencing
 CC of 5 different sizes of cDNA clone (see AAZ34570-74) suggests that
 CC at least 3 isoforms exist. The invention provides methods for the
 CC diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided for: suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 other;
 SQ
 Query Match 60.6%; Score 3080.2; DB 20; Length 5199;
 Best Local Similarity 79.8%; Pred. No. 0;
 Matches 4030; Conservative 0; Mismatches 753; Indels 268; Gaps 23;

Db 184 TAGCAAGGTAAAGTAAACAGAACCATGCTCAGTTTCAACACCTTTGGGAGGCTG 243
 Oy GATGCTGGGCGCATTAAGTGTGAGAGAAAGGCCAAGCATGACAGAGCTTCTAGGCTG 354
 Db 244 GATATCTGGGCGCATTAAGTGTGAGAGAAAGGCCAAGCATGATGACAGGTTCCATAGTTTA 303
 Oy 355 AAGCGATAGCGGAGATTATTAAGTGTGATGATCAAGCAGAGAACTTTTTTCCAACTGGG 414
 Db 304 AAGCCAAATATCTGATTCATTAAGTGTGATGATCAAGCAGAGAACTTTTTTCCAACTGGG 363
 Oy 415 TTACTGAGCTGCTTGAAGCAATATAGGCGGTACCGGAGATGAATAGATGAGAG 474
 Db 364 TTACTGAGCTGCTTGAAGCAATATAGGCGGTACCGGAGATGAATAGATGAGAG 423
 Oy 475 ATGATCAAGTGAATTTTTCATAGCATGAGAGCTTCAAACTGAAGTCAAGATAT 534
 Db 424 ATGATCAAGTGAATTTTTCATAGCATGAGAGCTTCAAACTGAAGTCAAGATAT 483
 Oy 535 CAGCTCCCTCCACACTTCCCTGTCATGAAACAGCAACCAAGTGTCTATTTCCAGTGA 594
 Db 484 CAGCTACCTCTGCACTTCCCTGTCATGAAACAGCAACCAAGTGTCTATTTCCAGTGA 543
 Oy 595 CCAGATTTGGTATAGAGAGGATTTGTCATGCAATGCGACACCTCAGGCTTGTCTGCTG 654
 Db 544 CCAGATTTGGTATAGAGAGGATTTGTCATGCAATGCGACACCTCAGGCTTGTCTGCTG 603
 Oy 655 CCAATGGGCTCATTTCCAGTTGTTGAATGTCACACCTTAGATCTTGTCTCTCA 714
 Db 604 CCAATGGGCTCATTTCCAGTTGTTGAATGTCACACCTTAGATCTTGTCTCTCA 663
 Oy 715 GCAGCAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 774
 Db 664 GCAGTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 723
 Oy 775 GCGCATCTCTGAGCCACATGCGCAAAAGTCTCTTCAAGCAGATGCTGCTGCTGCTGCT 834
 Db 724 GCTCATCTGAGCCACATGCGCAAAAGTCTCTTCAAGCAGATGCTGCTGCTGCTGCTGCT 783
 Oy 835 CAATTAACACTTAAGTTTCAAGAGGCAATCATTTGCTGCGAGCGGCTCTGCA 894
 Db 784 CAATTAACACTTAAGTTTCAAGAGGCAATCATTTGCTGCGAGCGGCTCTGCA 843
 Oy 895 GCAGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
 Db 844 GCAGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
 Oy 955 GACCAAACTATGAGTGAACACTTAAACAGTCCCGAGCAAGACTATTTCTGATCAATCA 1014
 Db 904 GACCAAACTATGAGTGAACACTTAAACAGTCCCGAGCAAGACTATTTCTGATCAATCA 963
 Oy 1015 AGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTTGAATGATCAAGATGA 1074
 Db 964 AGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTTGAATGATCAAGATGA 1023
 Oy 1075 AAATCTACTGAGAGAAATTAATCTTACTATGCACTTAATTAATTTGATGCTGCTGCT 1134
 Db 1024 AAATCTACTGAGAGAAATTAATCTTACTATGCACTTAATTAATTTGATGCTGCTGCTGCT 1083
 Oy 1135 CAGCCATGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194
 Db 1084 CAGCCATGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1143
 Oy 1195 TCCGAGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1254
 Db 1144 TCCGAGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203
 Oy 1255 CCGTGTGAGAGATGAGAGAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGAT 1311
 Db 1204 CCGTGTGAGAGATGAGAGAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGAT 1263
 Oy 1312 AAGAACCGGAGAACTTGTGAGAGAGAGTGTGAGAGAGCGCGGCAAGCGCTC 1371

QY	3529	GGGAGCTGATTCGTGATCCGGAAAAAAGAACCCAGGTGATGGTGAAAGGAACTCCAA	3588
Db	3271	GGTCAGCTGATTTTGTATCCGAAAAAAGAACCCAGGTGATGGTGAAAGGAACTCCAA	3330
QY	3589	GCTCGAGGAAAAAGCGCCAGATAGGGTGGTTTCCAGCAAAATTATGTCAACTTTAAGC	3648
Db	3331	GCACGTGGAAAAAGCGCCAGATAGGGTGGTTTCCAGCTAATTATGTAAGCTTTCTAAGC	3390
QY	3649	CCCGGAACAAAGCAAAATCACCCCACTGACTACCCAGACCGCATGACGAATCTAGCCTTAGC	3708
Db	3391	CTTGCGACGAGCAAAAACTCACTCCACAGAGCCACTTAAGTCAACGCACTTACCGCGAGTG	3450
QY	3709	TGCGAGGTGATCGGGAGTGTACGATTCACCGCCAGAAACGATGACGAATCTAGCCTTAGC	3768
Db	3451	TGCGAGGTGATGGAGTGTACGATTCACCGCGCAAGATGACGATGACTGGCTTTCAAC	3510
QY	3829	AGTGGCAAGTTGGGCTCTTCCCATCAAATTATGTAAAGCTGACCAAGACATGACCC	3888
Db	3571	AATGGAACATGGGGGCTCTTCCCATCAATTATGTGAAGCTGACCAAGACATGACCCA	3630
QY	3889	AGCCAGCAATGAATCATTATGTTGTCATCCGCCCTCAGGCTGTGAAGTCCCTCAAAGGA	3948
Db	3631	AGCCAGCAATGAATCATTATGTTGTCATCCGCCCTCAGGCTGTGAAGTCCCTCAAAGGA	3690
QY	3949	CCCACTATCCCATATCACTCTCCAGAGGAGATAGGAGATGACACCTTGATTCATGTGAC	4008
Db	3691	CCCACTATCCCATATCACTCTCCAGAGGAGATAGGAGATGACACCTTGATTCATGTGAC	3750
QY	4009	TTGAGCATATCACTACTGCTCTTGAGTAGAAGAACTCACCTCAGACAGATTAACT	4068
Db	3751	TTCCAGCATATCACTACTGCTCTTGAGTAGAAGAACTCACCTCAGACAGATTAACT	3810
QY	4069	CATTGACCTTAGTGTGATGTGATGCAAAATGTCTGATCACTGCGTGAAGGAGCAAGC	4128
Db	3811	CATTGACCTTAGTGTGATGTGATGCAAAATGTGTGATTAATCTGAGAGATAGGAGC	3870
QY	4129	--AAATTGCAAGACTGACAGGGTGGTGGGCTCTTTGGGGCTTCTCTAGTCACTGAC	4186
Db	3871	AAAAATTTCAAAAACACACAGGGTATGGGTCTTTTGTGGCTTCTCTAGTTCATCAAT	3930
QY	4187	TGAC--CGGCCCCGCTTCACACGGGGCTTTCATATGTTTAAATTAATTTTAAATGTG	4245
Db	3931	TGACCTTCCCCACCTTTGACACAGGGTCTTCAATGTTTAAATTAATTTTAAATTA	3990
QY	4246	TATTTTAGCTTTTAATAAAAATCTCATCATTAATCTCTTTGCTATTTTGGTTTACA	4305
Db	3991	TATTTTAGCTTTTAATAAAAAAATAATAATTAATGACTCTTTGCTATTTTGGTTTACA	4050
QY	4306	AAAAACCCCACTATCAAGAGTGTGCTGTGCGGAGCATTTAAATAGCTTCCGGGAGCTA	4365
Db	4051	AAAAAGACCCCACTATCAAGAGATG----CTGCATGTGCTATTTAAAAATTTTCCAAATGT	4105
QY	4366	CCGTAAACTGAGACTTGTCTGTAACCTTTGCGGTTTGTCCAGTGTCCAAACCAATTG--T	4424
Db	4106	CCATTAATCTGAGACTGTAGTATTTTTCATTTTGTCCAGTGTACCAACTAAATTTCT	4165
QY	4425	GTAAGTTGGGGCTGTTC--TGCCTTAGAGCAACAGAGAGATGGTGTACCTGTTT	4480
Db	4166	GCAGTTTGGGGCTTTTCCCTTACCATAGAGTGAAGAGAGTTGAGATCTCGTTT	4225
QY	4481	GAAATGTGATGTAGACTGAGGCTGCTATGGAAGGGGTATGCTGTGCTGTGACATC	4540
Db	4226	AAAGACGTATGATGAGCCCAATTAA--GCGAAGGTGATGTGCTTGTGTGTGTATC	4284
QY	4541	ACGTGTATC--TGTGCGCATGTACCATCTGTACCGAAGAGTAGCT--CTTCTCCATGAC	4598
Db	4285	AGCTGTATCCTTGTGTGAGATGTATATCATCTGTACATTAAGAAATAGTCTTTCATAGG	4344
QY	4599	TAAACCAACACCGGTGATAGTGTCTTCATC--TACTGCATTAATTTTACTTT--GCACAG	4655

Db	4345	CAAAAGCTATTACCTGTGACGATGCCTAATCATATTTGGCATTAATTTTATTTTGGAACAG	4404
Oy	4456	TGACCTTTGTACCACCTGAGAAGCAC-CCATGTTTTCCGTTTGGTCTCAGATGTAACCTAG	4714
Db	4405	TGACCTTTGTAGCCACATGAAAAGACTGTGTGTTTTTTGTCGATCAGATTTATCTCG	4464
Oy	4715	TTTGGCCCGGTTTTGTTTTTATTTTCAATCTGGCAGTCTTCACACATAACTAGTA	4774
Db	4465	TTGAGTTGGTGTTTTGTTTTGGGGTTTTTAATTTTGGTGTTTGATAGCATRAAATACGT	4524
Oy	4775	AGAGCCCACTGCCACGACGCGTTTACGATCATGATACCACCGTC-----TTAGTCCTG	4829
Db	4525	AGACAAC-AACACTGAGGTGTTAGCATCAACGATATCACAGTCTCTTTTAGTCTCG	4583
Oy	4830	TTACGCTAAC-TTTATTTCCAGTTCCTTTTATAGAA-----TATCTTGACAAGTAATCT	4883
Db	4584	TTACATCAAAGTTTTATTCAGTTACTTTTCATGGAATGACCTATTTTGAACAAGTAATTT	4643
Oy	4884	TCCTGACAAGAAAAAGATGTATAGAGTCCCCTGCATTAATTTTCCCCAGTGTTCATTT	4943
Db	4644	TCCTTGACAAGAAAAAGATGTATAGAGTCCCCTGCATTAATTT--CCAATGTTTACATTT	4702
Oy	4944	TTTAACTAGACTGTGGGGGTTTGTCACAGATTATATGAATGGCCCTCTGTGTCGGTGG	5003
Db	4703	TTTAACTAGAGACTGTGGAATTTCTACAGATTATATGAATGGAGCTATGTCGGTTTG	4762
Oy	5004	TGTGTTAACTTTGTGCTGTAGCTTAAGCCGTGT---CCTTAGATATTAGTTGGAAGTCG	5066
Db	4763	TGTGTTT-AGATATGCTGTACTGAAGCCCTGTTGTCTTTTAAACACTAGTGTGGAAGCTC	4821
Oy	5061	GGAAGAGAAATT 5071	
Db	4822	TCAATTAAAAAT 4832	

RESULT 8
AAS84763
ID AAS84763 standard; cDNA; 7435 BP.
XX
XX AAS84763;
AC
XX
DT 13-FEB-2002 (first entry)
DE
DE DNA encoding novel human diagnostic protein #20567.
XX
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss
XX
XX Homo sapiens.
OS
XX
PN WO200175067-A2.
PD
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG20576.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1; SEQ ID No 20567; 103pp; English.

Claim 1; SEQ ID No 20567; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. A864197-A894564 represent novel human
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/publ/published_pct_sequences.
CC
XX

Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 other;

Query Match 51.9%; Score 2638; DB 23; Length 7435;
Best Local Similarity 83.1%; Pred. No. 0;
Matches 3331; Conservative 0; Mismatches 615; Indels 64; Gaps 26;

QY 8 AGGAGAGTGGAGCGGCGCGGAGGCGCGACCTTGTTGCTCCGTAGTACGGCGGCTC 67
DB 22 AGAAGAGTGGATCGCGAGGAGGAGGAGGAGCTTGTTGCTCCGTAGTACGGCGGCTC 81
QY 68 GCAAGGAGGATCCCGAGCGGCGCTCCGCGAGCGCGGAGGAGGAGCGCGGCGGCGG 127
DB 82 GCGAGGAGGATCCCGAGCGGCGCTCCGCGAGCG-----GACGAGAGGCGGCGGCGG 130
QY 128 GATGTGTGTCGCGGCTCGGAGCTCGGCTCTCTGC-GGGGCGTCCGGGCTTCACTGAT 186
DB 131 GATGTGTGTCGCGGCTCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 190
QY 187 TTGTGTGAGGCGCGCGCGCGACCGCGCGGAGATGAGGCGTGCATGACGAAGTGA 246
DB 191 TTGTCTCTGCGGCGCGGAGCGCGCGCGGAGATGAGGCGTGCATGACGAAGTGA 250
QY 247 CGTAATAGAACCATGCTCATGTTTCCACACCTTGGGTGATGCTGATGCTGCGGCG 306
DB 251 AGTAACGAGAACCATGCTCATGTTTCCACACCTTGGGTGATGCTGATGCTGCGGCG 310
QY 307 ATAACTGTGAGGAGGAGGCGCGACGATGACGACAGTTCCTTAGCTGAGCGGATAGCG 366
DB 311 ATAACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 370
QY 367 GATTTATTAATCTGTGATCAAGGAGGAACTTTTTCCTTCAATCTGGTCTTACTGAGCT 426
DB 371 GATTTATTAATCTGTGATCAAGGAGGAACTTTTTCCTTCAATCTGGTCTTACTGAGCT 430
QY 427 GTCTTGAACAATATGCGGCTAGGCGGACATGATTAACGATGAGGAGATGATCAAGTG 486
DB 431 GTTTTGAACAATATGCGGCTAGGCGGACATGATTAACGATGAGGAGATGATCAAGTG 490
QY 487 GAATTTTCCATAGCCATGAAGCTTATCAAACTGAAGCTACAGAGATATAGCTCCCTTC 546
DB 491 GAATTTTCCATAGCCATGAAGCTTATCAAACTGAAGCTACAGAGATATAGCTCCCTTC 550
QY 547 ACATTTCCCTGTCTGAGAAAGAGCAACAGTGTATTTCCAGTCCACAGCATTTGGT 606
DB 551 GCACTTCCCTGTCTGAGAAAGAGCAACAGTGTATTTTTCAGGCGACAGCATTTGGT 610
QY 607 ATAGAGGAGATTGCTAGCATGCCACACTCAAGCTGTGCTCTGTGCAATGGGCTCC 666
DB 611 ATGGAGGAGATTGCTAGCATGCCACAGCTGTGCTCTGTGCTCAAGTGCATGGGATCC 670

QY 667 ATT-CAAGTGTGGAATGTCCTCCACCTTAGTATCTTCTGTCTCCAGAGAGTCC 725
DB 671 ATTCCAGTGTGGAATGTCCTCCACCTTAGTATCTTCTGTCTCCAGAGAGTCC 730
QY 726 TCCCTGGCTTACGGGGCTCCTCCGTCATACAGCCCTCCGCTGGCTTGGATCCTCG 785
DB 731 CCCCCTGGCTTACGGGGCTCCTCCGTCATACAGCCCTCCGCTGGCTTGGATCCTCG 790
QY 786 AGCCACATGGCCAAAGATTCTTCTTCAAGATCTGATCCAGGCTCACATTTAAAC 845
DB 791 AGCCACATGGCCAAAGATTCTTCTTCAAGATCTGATCCAGGCTCACATTTAAAC 850
QY 846 TAAGTTCAGAAAGCAATATTCATGATGTCGACAGGCCCCCTCC-AGCAGCAATGGG 904
DB 851 TAAATTCAGAAAGCAATATTCATGATGTCGACAGGCCCCCTCC-AGCAGCAATGGG 910
QY 905 CTGTGCTCAGTCTCAAGCTGAAATACA---GGCAGTATTCACAGCCACGACAAA 961
DB 911 CTGTGCTCAGTCTCAAGCTGAAATACA---GGCAGTATTCACAGCCACGACAAA 970
QY 962 C-TATGATGACACTT--AACAAGTCCCGAGCAAACTTCTCATGATCAATCAAGT 1017
DB 971 CCTAATGATGACACTTAAACAAGGCTCCCAAGCAAACTTCTTATGATGATCAAGT 1030
QY 1018 TTACCCAGGCTCAGTGGCTTCAATATGATCTTCTGACATGATCAAGTGAAGAA 1077
DB 1031 TTACCCAGGCTCAGTGGCTTCAATATGATCTTCTGACATGATCAAGTGAAGAA 1090
QY 1078 CTCACTGAGAAATTTATCTAGTATGACCTTAATGATGATGATGATGATGATGATG 1137
DB 1091 CTCACTGAGAAATTTATCTAGTATGACCTTCAATGATGATGATGATGATGATGATG 1150
QY 1138 CCAGTGGCGCGCGGCTCGCTCCAGAAATATATCTTCTTCTGAGAGAGTTCGCTCC 1197
DB 1151 CCAGTGGCGCGCGGCTCGCTCCAGAAATATATCTTCTTCTGAGAGAGTTCGCTCC 1210
QY 1198 GCGAGTGGATGTCCTCATAGCTCTTCTTCTGATGATGAGGCTGCTGAGAGCGG 1257
DB 1211 GCGAGTGGATGTCCTCATAGCTCTTCTTCTGATGATGAGGCTGCTGAGAGCGG 1270
QY 1258 TCGTCAGAGTGAAGCGAGC---CAGAGAAATCTGCTGTGACATTTGAAGTAA 1314
DB 1271 GTTTTGAAGTGAAGCGAGCAACAATTTGAAGAAATTTACCTTGAAGTAA 1330
QY 1315 AAGCGGAGAACTTCAGCGAGGAGTGTGAGTGGAGAGCGCGCCAGAGCTCTTG 1374
DB 1331 AAGCGGAGAACTTCAGCGAGGAGTGTGAGTGGAGAGCGCGCGCCAGAGCTCTTG 1390
QY 1375 GAGCAGAGCGCAAGAGCAGAGCGGTTGCTTCACTGAGCGCGCGAGAGAGG 1434
DB 1391 GAGCAGAGCGCAAGAGCAGAGCGGTTGCTTCACTGAGCGCGCGAGAGAGG 1450
QY 1435 AAGAGGCGGAGCGCGCAGAGCGAGGAGCGGAGCGGAGCTGAGAGCGAGCTG 1494
DB 1451 AAGAGGCGGAGCGCGCAGAGCGAGGAGCGGAGCGGAGCTGAGAGCGAGCTG 1510
QY 1495 GAGAGAGCGGAGCTGAGCGGCGACGAGAGAGAGAGAGAGAGATCGAGAG 1554
DB 1511 GAGAGAGCGGAGCTGAGCGGCGACGAGAGAGAGAGAGAGAGATCGAGAG 1570
QY 1555 CGCAGGCGCGGAGAGCGGAGCT--GAGAAAGCGAGCAACTTGAATGGAGAGGAA 1613
DB 1571 CGCAGGCGCGGAGAGCGGAGCT--GAGAAAGCGAGCAACTTGAATGGAGAGG 1630
QY 1614 GAGAGAGAACTCTGATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1673
DB 1631 AGAGTCCGAGTCTAATTAATCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1690
QY 1674 AAGAGAGAGCTCTGAGATTGAGATTGAAGCTTGAATGACAAAGAGATGAGCTTGA 1733
DB 1691 AAGAGAGAGCTCTGAGATTGAGATTGAAGCTTGAATGAGAGAGATGAGCTTGA 1750
QY 1734 AGGAAACTTCAGGATATCAGGTGTGAGCTGAGCAACCGAGGCAAGAAATGAGAGC 1793

```

1751 AGGAAATCTAGAGATATGATGTGATGACACCAAGCAAAAGCAAAATGAGGAC 1810
1794 GAAAGATCTAGAGATTAAGATTGCTG--AAATCACCCACTTACAGCAGAGT--GC 1849
1811 AAACAAATCTAGAGATGTAAGATTGCGGAAATCAACCCATCTACAGACTAATTTACA 1870
1850 AGAATCTCAGCAAA--TGCTTGAAGAATTATTCAGAGAAACAGATCACTAGTACC 1906
1871 TGGAACTCTGAGCAAAATGCTTTGGAAGAACTTATTCAGAAAAACGATATCAATGACC 1930
1907 AGTTAAAAAGTCAGAGAACAGATTGATGAGACTGCTTTTACCCTCAAAAG-A 1965
1931 AATTAAAAAGTCAGAGAACAGATTGATGAGACTGCTTTTACCACTTAAAAAGAA 1990
1966 GCCTTGAAGCAAAAGAGAGTGGCGCCGAGCAGCTCCGAGAGAGTGAAGAGGAGAG 2025
1991 GCCTTGAAGCAAAAGAGAGTGGCGCCGAGCAGCTCCGAGAGAGTGAAGAGGAGAG 2050
2026 AGAGAGCAGGTCAAAAGCTGAGAGATTTGATTTTCAACCAACAGCTGAAGAACTG 2085
2051 AAGAAATCTAGATCAAAACTACAGAGATTGATATTTTCAATATCAGCTGAAGAACTA 2110
2086 AGAGAGTATATGCAAAACGCAATTCAGAGACAGAGTCCCTGAGGAGCGGACTG 2145
2111 AGAATAATACACATTAAGCAAACTCCAGAACCAAAAGTCATGAGAGCTGAACGACTG 2170
2146 AACGAGAAAGAGAGAGAGAGAGAGTGGAGTTAGAGAGCAAAAGAGAGCGCTGAG 2205
2171 AAACAGAAACAGAGAGAGAGAGAGATGATTAAGAAATTAAGAAACAAAGAGAGAGAG 2230
2206 AGAGAGTTCAGAGAAAGAGAGAGAGAGATGAGATGAGAGATGAGAGAGAG--GAGCAG 2262
2231 AGAGAGCTCAGAGAAAGAGAGAGAGAGAGTGGCTGAGAGATGAGAGAGAGAGAGAT 2290
2263 CCAAGCCCCGAGAAACCCACAGAGAGAGAGAGAGTGAAGAGGAGAGAGAGAGAGAG 2322
2291 CAGAGAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2350
2323 AAGGA--GGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2381
2351 AAGGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2410
2382 TCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2441
2411 TCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2470
2442 CCGGCTTACATTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2501
2471 TCCACTTACATTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2530
2502 CTTTGAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2555
2531 CTTTGAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2590
2556 -----GGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2606
2591 AGGGAGATGGGTGGATGAAGCCAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2650
2607 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2666
2651 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2710
2667 CACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2726
2711 CGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2770
2727 TGAGAGCCCTGCTCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2786
2771 TGAGAGCCCTGCTCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2830
2787 GCGAGACTTCAGTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2846

```

```

2831 GCGGACTTCAGCTCCACAGTGGCCACAGCAGATGAGAGAAACAGAGAGAGAGAGAGAG 2890
2847 GAGATACGTGGGCGGCTCAG--CTTCTCTGACCGTACCTAGTGTGCGCACTTACGAGAG 2905
2891 GAGATACGTGGGCGGCTCAG--CTTCTCTGACCGTACCTAGTGTGCGCACTTACGAGAG 2950
2906 GATCAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2965
2951 GATCAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3010
2966 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3024
3011 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3070
3025 AACCACTTAAATTTTAAACAAAGTGAAGTCAACCGTTCTGAGAGAGAGAGAGAGAGAG 3084
3071 AACCACTTAAATTTTAAACAAAGTGAAGTCAACCGTTCTGAGAGAGAGAGAGAGAGAG 3130
3085 TGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3144
3131 TGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3190
3145 TCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3203
3191 TCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3250
3204 TCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3262
3251 TCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3310
3263 CCATGATACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3318
3311 CCATGATACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3370
3319 GTGATTTGAGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3377
3371 GTGATTTGAGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3430
3378 CGAGTCTTCTCTCTTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3437
3431 CGAGTCTTCTCTCTTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3490
3438 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3496
3491 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3550
3497 CTACGTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3556
3551 CCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3610
3557 A--CCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3615
3611 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3670
3616 TGGTTTCCAGCAATTTATGTAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3675
3671 TGGTTTCCAGCAATTTATGTAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3730
3676 GAGCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3735
3731 GAGCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3790
3736 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3794
3791 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3850
3795 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3854
3851 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3910
3855 CAATTATGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3914
3911 CAATTATGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3970

```


Qy 3915 ATCCCCCCTCAGGCTTGAAAGTCCCTCAAGAAGACCCACTATCCCATATC 3964
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3971 ATCTCTTGATATGTGACCCCACTGAAGAAGCGACAAGATACATC 4020

RESULT 9	
AA163825	
ID	AA163825 standard; cDNA; 3466 BP.

AC AA163825;

DT 22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 33.

KW Human, antiatheritic; antiapneumatic; antiproliferative; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; vitricide
 KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
 KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
 KW antiinflammatory; antitumor; vulnerary; anticonvulsant; antibacterial,
 KW antiparasitic; cardiatic; gene therapy; cancer; immune disorder;
 KW cardiovascular disorder; neurological disease; infection; human; ss.
 XX
 XX Homo sapiens.

Os Homo sapiens.

PN W0200155308-A2.

02-AUG-2001 PD

PF 17-JAN-2001; 2001WO-US01309.

PR	31-JAN-2000;	2000US-0179065;
PR	04-FEB-2000;	2000US-0180628;
PR	24-FEB-2000;	2000US-0184664;
PR	02-MAR-2000;	2000US-0186350;
PR	16-MAR-2000;	2000US-0189874;
PR	17-MAR-2000;	2000US-0190076;
PR	18-APR-2000;	2000US-0198123;
PR	19-MAY-2000;	2000US-0205515;
PR	07-JUN-2000;	2000US-0209657;
PR	28-JUN-2000;	2000US-0214686;
PR	30-JUN-2000;	2000US-0215135;
PR	07-JUL-2000;	2000US-0216647;
PR	07-JUL-2000;	2000US-0216880;
PR	11-JUL-2000;	2000US-0217487;
PR	11-JUL-2000;	2000US-0217496;
PR	14-JUL-2000;	2000US-0218290;
PR	26-JUL-2000;	2000US-0220963;
PR	26-JUL-2000;	2000US-0220964;
PR	14-AUG-2000;	2000US-0224518;
PR	14-AUG-2000;	2000US-0224519;
PR	14-AUG-2000;	2000US-0225213;
PR	14-AUG-2000;	2000US-0225214;
PR	14-AUG-2000;	2000US-0225266;
PR	14-AUG-2000;	2000US-0225267;
PR	14-AUG-2000;	2000US-0225568;
PR	14-AUG-2000;	2000US-0225570;
PR	14-AUG-2000;	2000US-0225547;
PR	14-AUG-2000;	2000US-0225757;
PR	14-AUG-2000;	2000US-0225758;
PR	14-AUG-2000;	2000US-0225759;
PR	18-AUG-2000;	2000US-0226279;
PR	22-AUG-2000;	2000US-0226681;
PR	22-AUG-2000;	2000US-0226688;
PR	22-AUG-2000;	2000US-0227182;
PR	23-AUG-2000;	2000US-0227009;
PR	30-AUG-2000;	2000US-0228924;
PR	01-SEP-2000;	2000US-0229287;
PR	01-SEP-2000;	2000US-0229343;
PR	01-SEP-2000;	2000US-0229344;
PR	01-SEP-2000;	2000US-0229345;
PR	05-SEP-2000;	2000US-0229509;

[illegible]

PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249219.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0255719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX			
PA	(HUMA-)	HUMAN GENOME SCI INC.	
PI	Rosen CA,	Barash SC, Ruben SM;	
XX			
DR	WPI;	2001-488781/53.	
XX			
PT	P-PSDB;	AAM43519.	
XX			
PS	Claim 1;	SEQ ID NO 33; 664bp + Sequence Listing; English.	
XX			
CC	The invention relates to human polynucleotides (AA163803-AA164012) and		
CC	the encoded proteins (AAM434497-AAM43660) useful for preventing, treating		
CC	or ameliorating medical conditions e.g. by protein or gene therapy. The		
CC	genes were isolated from a range of human tissues disclosed in the		
CC	specification. The nucleic acids, proteins, antibodies and (ant)agonists		
CC	are useful in the diagnosis, treatment and prevention of: (a) cancer,		
CC	e.g. breast and ovarian cancer and other cancers of the adrenal gland,		
CC	bone, bone marrow, breast, gastrointestinal tract, liver, lung, or		
CC	urogenital; (b) immune disorders e.g. Addison's disease, allergies,		
CC	autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,		
CC	Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative		
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;		
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and		
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal		
CC	and parasitic infections.		
CC	Note: The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
XX			
XX	Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 10 other;		
QY	Query Match	43.1%; Score 2188.8; DB 22; Length 3466;	
	Best Local Similarity	84.7%; Pred. No. 0;	
	Matches 2525; Conservative	0; Mismatches 427; Indels 30; Gaps 5	
QY	7	GAGGAGGAGTGGACGGCGCGGGAGGGCCGACGTTGGTGTCTCGTAGTACGGCGCT	66
DB	35	GAGAAAGTGGAGGGCCGACGAGGGAGGAGACGTAGTGGTGTCTCGTAGTACGGCGCT	94
QY	67	CGCAAGGAGCATCTCCGAGCGCGCTCCGGAGAGCGCCGGGAGGACGAGCGCGCGGGCG	126
DB	95	CGCAGGAAATAATCCGAGCGGCTCCGGAGGCG-----GACGAGAGGCGGGGCG	143
QY	127	GGAGTGTGTGCGCGGCTGGCGACTCGGCGTTCTCTGC-GCGGCGTGGGCTGCACCTGA	185
DB	144	GGAGTGTGTGCGGGGCTGGCGCTCTCTGCTCTCCACGCGGCGCGGAGCGGACACTGA	203
QY	186	TTTGTGTGAGGGCGCGCGCGCACCGCCGAGATGAGGCGGTGATCGACGAAGGTGA	245

Db	204	TTTGTCCCTGGGGGGCCAGGCGGAGCCCGCCGGAGATGAGGGCGTGCATTAGCAAGGTA	263
Qy	246	ACGTAATAGAACCATGGCTCAGTTTCCACACCTTTCCGTGGTAGCTTGATGTCGGGC	305
Db	264	AAGTAAAGAAACCATGGCTCAGTTTCCAACACCTTTGGTGGCACTCGATATCTGGGC	323
Qy	306	CATTACTGTGGAGAAAAGGGCCCAAGATACACAGCAGTTCTTACCCTGAAGCCGATAGC	365
Db	324	CATTAACGTGTAGAGAAAAGCCGAAGCATATAGCAGTTCCATTAAAGCCAAATATC	383
Qy	366	GGGATTTATTAAGTGTGATATCAGCGAGAACTTTTTTCCATGTGGGTTACCTCAGCC	425
Db	384	TGGATTAATTAAGTGTGATCAGCTGAAACTTTTTTTTCAATCTGGGTTACCTCAACC	443
Qy	426	TGTTTAGACAATAATATGGCGCTAGCGGACATGAATAACGTGAGAGATGATCAAGT	485
Db	444	TGTTTAGACACAGATATGGGCACTAGCTACATGAATAATATGAGAAATGATATCAAGT	503
Qy	486	GGAAATTTTCCATAGCCCATGAAGTTTACAACTGAAGCTACAGAAATATCAGTCCCTC	545
Db	504	GGAAATTTTCCATAGCTATGAATCTTACAACTGAAGCTACAGAAATATCAGTCCCTC	563
Qy	546	CACACTTCCCCCTGTATGTAAGACAGAACCCAGTGGTATTTCCATGACACAGCATTTGG	605
Db	564	TGCATTTCCCCCTGTATGTAAGACAGAACCCAGTGGTATTTCCATGACACAGCATTTGG	623
Qy	606	TATAGAGGAGATTGCTAGCATGTCACACACTCAAGCTGTGCTCCTGTCCMATGGGCTC	665
Db	624	TATGGAGGATATGGCCAGCATGTCACCGCTTACAGTGTGCTCCAGTGCCMATGGGATC	683
Qy	666	CATTCCAGTTGTTGGAAATGTCCTCACCCCTTAAATCTTGTCTCTCCAGCAGCAGTGGC	725
Db	684	CATTCCAGTTGTTGGAAATGTCCTCAACCCCTAGATCTTCTGTTCCACAGCAGCAGTGGC	743
Qy	726	TCCCGTGGCTAAAGGGGCTCTCCGCTCATPAGGCTCGCTGGCTTGGGCAATCTGC	785
Db	744	CCCCGTGGCTAAAGGGGCTCCCCCTGTATPACAACCTCGCTGGATTTGCTCATCTGC	803
Qy	786	AGCCACATGGCCAAAGATCTTCTTTCAGCAGATCTGGTCCAGGGTCACAAATTAACAC	845
Db	804	AGCCACATGGCCAAAGATCTTCTTAAATGATCTGGTCCAGGGTCACAAATTAACAC	863
Qy	846	TAACTTACAGAAAGCACATCATTTGATGTCCGACGCGCCCTCCAGCAGCAGAAATGGCC	905
Db	864	TAAATTTCAAAAAGGCACAGTCATTATATGTGGCAGTGTCCACACAGTGGCAGATGGGC	923
Qy	906	TGTGCCCTACATCAAGGCTGAATAACAGGAGATTATCAACAGCCACGACAAACAT	965
Db	924	TGTTCCTCAGTCATCAAGACTGAATAACAGGACATTTCAATACATCAGCAAAACAT	983
Qy	966	GAGTGGACACTTAAC-----AAGTCCCCAGGCAAGACATATTTCATGCAATC	1013
Db	984	GAGTGGACACTTAACAGGTTCCGTTTAGTGTCCCAAGCAAAACATATTTCATGCAATC	1043
Qy	1014	AAGTTTACCCAGGCTCAGCTGCTCAATATGAATCTTTGTGACATTTGATCAAGATGG	1073
Db	1044	AAGTTTACCCAGGCTCAGCTGCTCAATATGAATCTTTGTGACATTTGATCAAGATGG	1103
Qy	1074	AAAACCTCACTGCAAGAAATTTATCTTACTATGCACTTAATTGATGTTGGCATGTCTGG	1133
Db	1104	AAAACCTTACAGCAGAGAAATTTATCTTGGCAATGCACTCATTTGATGTAATGTCTGG	1163
Qy	1134	TCAGCAGCTGGCGCGCTCTGCTGCCAAGATACATCCCTCTTCTCAGAAAGTTCTG	1193
Db	1164	CCAAACCACTGCCACTGTCTCTGCTCCAGAAATACATTCACACTTCTTTTAAAGAGTTCC	1223
Qy	1194	CTCCGCGAGTGGGATGTCCGCTCAATAGCTCTTCTTGTGATTCAGAGCTGCTTGAGA	1253
Db	1224	ATCTGGCAGTGGTATATCTGTCAATAGCTCAACATCTGTAGATCAGAGGCTACAGAGGA	1283
Qy	1254	GCCTGCTCAGAGATGAGAGACAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGA	1310

Db 1284 ACCAGTTTGAAGATGACACACAACTATTAGAAAAAATTACTGTATCGTTTGAGA 1343
 Qy 1311 TAAGAGCGGGAGAACTTTCAGCGAGGCACTGTGAGCTGTGAGAACGCCGCCAACCGCT 1370
 Db 1344 TAAGAGCGGGAGAACTTTCAGCGAGGCACTGTGAGAACGCCGCCAACCGCT 1403
 Qy 1371 CTGAGACAGCAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGAGACCGCGCAGAGAGA 1430
 Db 1404 CTGAGACAGCAGCGCAAGAGCAGAGCGGCTGGCCAGCTGTGAGCGGCGAGAGAGA 1463
 Qy 1431 GAGAGAGAGCGGAGCGCCAGAGAGCAGAGCGCCAGCGCAGCTGTGAGAGAGA 1490
 Db 1464 GAGAGAGAGCGTGTGAGCGCCAGAGAGCGCCAGAGAGCGCAGCTGTGAGAGAGA 1523
 Qy 1491 GCTGAGAGAGCGGAGCGGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1550
 Db 1524 ACTGAGAGAGCGGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1583
 Qy 1551 GAGGCGGAGCGGAGCGGAG 1610
 Db 1584 GAGGCGGAGCGGAGCGGAG 1643
 Qy 1611 CCGGAG 1670
 Db 1644 TCGAAG 1703
 Qy 1671 GCGAG 1730
 Db 1704 AGCAAG 1763
 Qy 1731 AGAAG 1790
 Db 1764 AGAAG 1823
 Qy 1791 CAGGAG 1850
 Db 1824 CAGGAG 1883
 Qy 1851 GGAATCTCAGCAATGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1910
 Db 1884 GGAATCTCAGCAATGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943
 Qy 1911 AAAAAG 1970
 Db 1944 AAAAAG 2003
 Qy 1971 GGAAG 2030
 Db 2004 AGAAG 2063
 Qy 2031 GACGAG 2090
 Db 2064 AACTAG 2123
 Qy 2091 GATGAG 2150
 Db 2124 AATGAG 2183
 Qy 2151 GAAAG 2210
 Db 2184 GAAAG 2243
 Qy 2211 AGTTGAG 2267
 Db 2244 AGCTGAG 2303
 Qy 2268 CCGGAG 2327
 Db 2304 ACCAG 2363
 Qy 2328 GCGGAG 2387
 Db 2364 TGGGAG 2423

Qy 2388 TCAGAGCAGACTAAGCTGGCCACCAGGACCCCTGCTTACACAGAGAGAGAGAGAGAGAGAG 2447
 Db 2424 CCAAGAGCAGACTAAGCAGCTGTCCAGGACCCCTGCTTACACAGAGAGAGAGAGAGAGAGAG 2483
 Qy 2448 TACATTTCTGACAG 2507
 Db 2484 TACATTTCTGACAG 2543
 Qy 2508 ATCCAG 2567
 Db 2544 ATCCAG 2603
 Qy 2568 CCAG 2627
 Db 2604 CCAAG 2663
 Qy 2628 TGCAAACTATGCAAG 2687
 Db 2664 TGCAAACTATGCAAG 2723
 Qy 2688 CGATGTGACATCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2747
 Db 2724 TGATTCAGATCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2783
 Qy 2748 AGTGAAGCTCTTCTGAG 2807
 Db 2784 AGTGAAGCTCTTCTGAG 2843
 Qy 2808 GCCAG 2867
 Db 2844 GCCAG 2903
 Qy 2868 TTCTCTGAG 2927
 Db 2904 CTCTCTGAG 2963
 Qy 2928 AGCAG 2969
 Db 2964 GGCAG 3005

RESULT 10
 ABK43498
 ID ABK43498 standard, cDNA; 3319 BP.
 XX
 AC ABK43498;
 DT 05-JUN-2002 (first entry)
 XX
 DE DNA encoding novel central nervous system protein #78.
 XX
 KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;
 KW myocardial infarction; wound healing; cell proliferation; skin aging;
 KW food additive; food preservative; gene therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200155318-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01332.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234297.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254078.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-581633/65.
DR P-PSDB; AAU87168.
XX
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
XX food additives or preservatives -
XX

PS Claim 1, SEQ ID No 88, 837bp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnoses of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocatalal
CC infection. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein.

Query Match 42.1%; Score 2140.8; DB 23; Length 3119;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 2447; Conservative 0; Mismatches 412; Indels 19; Gaps 4;

QY 111 GCGAGCGGGCGGGCGGATGCTGCGGGCGGATCGGCGCTTCTGCGC-GCGGC 169
DB 10 GACAGAGGCGGCGGCGGATGCTGCGGGCGGCTGCGCTTCTTCCACCGGC 69
QY 170 GTCGGGCTCACTGATTTGTGTAGGGGCGCGCGCGCACCGCGCGAGATGAGGG 229
DB 70 GCGTGAAGCGCATGATTTGTCTCGGGCGCGAGCGCGGACCGCGGAGATGAGGG 129
QY 230 TCGATCAGCAGGTGAACGTAATGAACCATGCTCATGTTCCACACCTTTGGTGTA 289
DB 130 TCGATTAAGCAAGGTAAAGTAACGAAACCATGCTCATGTTCCACACCTTTGGTGTA 189
QY 290 GCGTGAAGCTGGGGCATTAAGTGGAGGAAAGGCGAAGCATGACGAGCTTCTTA 349
DB 190 GCGTGAATCTGGCCATTAAGTGGAGGAAAGGCGAAGCATGATCAGACCTTCTTA 249
QY 350 GCGTGAAGCGGATGCGGATTTATTAATGATCAAGGAGAACTTTTTTTCAT 409
DB 250 GTTTAAAGCAATATCTGATCTTACTGATCAAGTAACTTTTTTTCAT 309
QY 410 CTGGGTTACTGACCTCTTGTAGCAAAATATGGGCGCTAGGAGCATGAATGAAGATG 469
DB 310 CTGGGTTACTGACCTCTTGTAGCAAAATATGGGCGCTAGGAGCATGAATGAAGATG 369
QY 470 GAAGGATGATCAAGTGAATTTTCCATAGCATGAAGCTTATCAAGTGAAGCTCAAG 529
DB 370 GAAGGATGATCAAGTGAATTTTCCATAGCATGAAGCTTATCAAGTGAAGCTCAAG 429
QY 530 GATATGAGTCCCTCTCAACCTTCCCTGTCAATGAAGCAAGCAACGATGGCTATTTC 589
DB 430 GATATGAGTCCCTCTCAACCTTCCCTGTCAATGAAGCAAGCAACGATGGCTATTTC 489
QY 590 GTGACACGAGATTGGATGAGGAGATTGCTAGCATGCCACACCTCAAGCTGGTCTC 649
DB 490 GTGACACGAGATTGGATGAGGAGATTGCTAGCATGCCACACCTCAAGCTGGTCTC 549
QY 650 CTGTGCAATGAGGCTCCATTCAGATTGTGAATGTCTCCACCTTATGATCTTCTGCTC 709
DB 550 CAGTGGCAATGAGGATTCATTCAGTTGTGAATGTCTCCACCTTATGATCTTCTGCTC 609
QY 710 CTTCAGAGAGAGTCTCTCCCTGGCTAGCGGGCTCTCCCTGATCAAGCTCTGCTG 769
DB 610 CCAAGAGAGTGTGCTCCCTGGCTAGCGGGCTCTCCCTGATCAAGCTCTGCTG 669

QY 770 CGTTTGGCATCTCTGACGACACATGGCCAAAGATTCTTCTTACAGAGATCTGTCTCAG 829
DB 670 CATTTGGCATCTCTGACGACACATGGCCAAAGATTCTTCTTACAGAGATCTGTCTCAG 729
QY 830 GGTCACAATTAACACTTAAGTTACAGAGGACATATTCATGATGTCCAGGCGCCCTC 889
DB 730 GGTCACAATTAACACTTAAGTTACAGAGGACATATTCATGATGTCCAGGCGCCCTC 789
QY 890 CAGAGAGGAGATGGGCTGTGCTCAGTCAAGAGGCTGAATACAGGACAGTATTACACA 949
DB 790 CAGGCGAGAGATGGGCTGTGCTCAGTCAAGAGGCTGAATACAGGACAGTATTACACA 849
QY 950 GCCAGCAAAACTATGATGAGACCTTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 997
DB 850 GTCTAGCAAAACTATGATGAGACCTTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 909
QY 998 CTATTTCATGCAATCAAGTTTACCCAGGCTCAGGCTTCAATATGGAATCTTCTCTG 1057
DB 910 CTATTTCATGCAATCAAGTTTACCCAGGCTCAGGCTTCAATATGGAATCTTCTCTG 969
QY 1058 ACATTGATCAAGATGAGAACTCACTGACAGAGATTTATCTAGCTATGACCTTAATTT 1117
DB 970 ACATTGATCAAGATGAGAACTCACTGACAGAGATTTATCTAGCTATGACCTTAATTT 1029
QY 1118 ATGTTGCCATGTCTGTGACGACCTGCGCCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1177
DB 1030 ATGTTGCCATGTCTGTGACGACCTGCGCCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1089
QY 1178 CTTTCAGAGAGTGTGCTGCGGCGAGTGGAGTGTCCGTCATTAAGCTCTTCTTGTGATC 1237
DB 1090 CTTTCAGAGAGTGTGCTGCGGCGAGTGGAGTGTCCGTCATTAAGCTCTTCTTGTGATC 1149
QY 1238 AGAGCTGCTCTGAGAGACCTGCTCAGAGATGAGCAGCAGC---CAGAGAGAACTGC 1294
DB 1150 AGAGCTGCTCTGAGAGACCTGCTCAGAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1209
QY 1255 CTGTGACATTTGAAGATTAAGAGCGGAGAACTTCAGCGAGGACAGTGTGAGCTGAG 1354
DB 1210 CTGTGACATTTGAAGATTAAGAGCGGAGAACTTCAGCGAGGACAGTGTGAGCTGAG 1269
QY 1355 AGCGCGCGCAAGCGCTCTTGGAGCAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGG 1414
DB 1270 AGCGCGCGCAAGCGCTCTTGGAGCAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGG 1329
QY 1415 AGCGCGCGCAAGCGCTCTTGGAGCAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGG 1474
DB 1330 AGCGCGCGCAAGCGCTCTTGGAGCAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGG 1389
QY 1475 TGAAGCTGGAAGCAGCTGGAAGAGAGCGGAGCTGAGCGGCGAGCGAGCGAGGAGAGA 1534
DB 1390 TGAAGCTGGAAGCAGCTGGAAGAGAGCGGAGCTGAGCGGCGAGCGAGCGAGGAGAGA 1449
QY 1535 GAGAGAAAGAGATGAGAGGCGCGAGGCGCAAAACGGGAACTTGAAGAGCGAGCAAC 1594
DB 1450 GAGAGAAAGAGATGAGAGGCGCGAGGCGCAAAACGGGAACTTGAAGAGCGAGCAAC 1509
QY 1555 TTGAATGGGAACGAAACCGAGAGAGAACTCTGGAATCAGAGAAACAAAGAGAGAGAG 1654
DB 1510 TTGAATGGGAACGAAACCGAGAGAGAACTCTGGAATCAGAGAAACAAAGAGAGAG 1569
QY 1655 GCACGCTGCTCTGAGAGCAGAGAGAGAGCTTGAAGTTTGAAGTTTGAAGCTCTGAATG 1714
DB 1570 GCACGCTGCTCTGAGAGCAGAGAGAGAGCTTGAAGTTTGAAGTTTGAAGCTCTGAATG 1629
QY 1715 ACAAAGAGATCACTTGAAGAGAACTTCAAGATATCAGGTGTGCTGCACTGCAACCCAGA 1774
DB 1630 ACAAAGAGATCACTTGAAGAGAACTTCAAGATATCAGGTGTGCTGCACTGCAACCCAGA 1689
QY 1775 GCGAAGAAATTTGAAGAGAGAGAGCAAGCTGAGAGCTTGAAGTTTGAAGTTTGAAGTT 1834
DB 1690 GCGAAGAAATTTGAAGAGAGAGAGCAAGCTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 1749

QY	1835	TACGACGCGAGTTGACAGGATCTCACCAATGCTTGAAGAATTATTCAGAGAAACGA	1894
Db	1750	TACGACCACAATTACAGGATCTCAGCAATGCTTGAAGAATTATTCAGAGAAACGA	1809
QY	1895	TACTCAGTGAACAGTTTAAAACAAGTCCAGCAGAACAGTTTGCATAGAGACTCGCTTTA	1954
Db	1810	TACTCAATGACCAATTTAAAACAAGTTCAGAGAACAGTTTGCACAGAGATTCACTTGTTA	1869
QY	1955	CCCTCAAAAAGACCTTGGAAAGCAAAGAGCTGGCCCGGAGACGCTCCGGAGACGCTGG	2014
Db	1870	CACTTAAAAAGCGCTTAGAAGCAAAAAGAACTAGCTTCGAGCAGCACTACGAGACCAACTGG	1929
QY	2015	ACGAGGTGGAGAGAGACCCAGGCTCAAGGTGAGAGATGATGTTTCAACAAACGAC	2074
Db	1930	ATGAAGTGGAGAAAAGAACTAGATCAAAAATACAGAGATTGATTTTCAATATCTACG	1989
QY	2075	TGAAGAACTGAGAGAGATCATATGCAAAAACAGCAATCCAGAAAGCAGAGTCCCTGAGG	2134
Db	1990	TGAAGAACTTAGAGAAATATCAATATGAAACAACTCCAGAAAGCAAAAGTCATGAGGAG	2049
QY	2135	CAGCGCACTGAAAGCAGAAAAGACGAGAGAGACAGAGCCTGGAATTAGAGAAACAAAAG	2194
Db	2050	CTGAACCACTGAAAACAGAAAAGAAACAAGAACTATAGAAATTAGAAAACAAAAG	2109
QY	2195	AAGCGCTCAGAGCAGGATCAGAAAAGGGAACAAGATGGCTGGAGCATGTGCGACGAG	2254
Db	2110	AAGAAAGCCCAAGACGAGCTCAGAAAAGGGAACAAGATGGCTGGAGCATGTGCGACGAG	2169
QY	2255	AG-- --GAGCAGACACGCCCCCGGAAAACCCACAGAGAGACAGACTGAAGAGGCAAGCA	2311
Db	2170	AGGACGAGCATCAGAGACCAAGAAAACTCCACGAACAGAAAAACTGAAAAGGAGAGAG	2229
QY	2312	GTGTCAAGAAAGAGAGCGGAAGAGAGACCAAGCCGGAATTCAGACAAACAGAGTC	2371
Db	2230	GTGTCAAAAAGAAAGATGCGGAGAAAAGGCAAAACAGAAACCAACAGACAAGCTGGGTC	2289
QY	2372	GGCTTTTCCATCCGCATCAGAGACCAAGTAAAGTGGCCACCAAGCACCCTGCTTACCA	2431
Db	2290	GGCTTTTCCATCAACACACAAAGCAACAGCTAAGCAGCTGTCCAGGACCCCTGGTCACTG	2349
QY	2432	CAGAGAAAGCCCGCTTACCATTTCTGCAACAGAGAGTGTAAAAAGTGATATTATTCAG	2491
Db	2350	CAGAAAAAGGTCCACTTACATTTCTGCAACAGAAAAATGTAAGAGTGTATTTCACCGG	2409
QY	2492	CGCTGTACCCCTTTGAAATCCAGAAAGTCAACGATGATCAACATCCAGCCAGAGATATAG	2551
Db	2410	CACGTACCCCTTTGAATTCAGAAAGCCATGATGAATTCATATCCAGCCAGAGACATAG	2469
QY	2552	TCATGTGTGATGAAGCCAGACTGAGAGCCAGAGATGGCTTGAAGAGAGCTGAAGAGGA	2611
Db	2470	TCATGTGTGATGAAGCCAGAACTGGAAGAACCCCGCTGGCTTGGAGAGATTTAAAAAGGA	2529
QY	2612	AGACGGGATGGTTCCTCGCAAACTATGACAGAAAAGTTCCAGAAAATGAGTTCACATC	2671
Db	2530	AGACAGGATGGTTCCTCGCAAACTATGACAGAAAATCCAGAAAATGAGTTCCTCGCTC	2589
QY	2672	CAGCAAAACAGTGAACCGATCTGACATCTGCCCCCTCCCAACCTGAGCTCTGCGGTGAG	2731
Db	2590	CAGTGAAAACAGTGAATCGATTCAACAATCTGCCCCCTCCCAACCTGAGCTCTGCGGTGAG	2649
QY	2732	CCCTGCTCTTTGCGCAGTACCTCTTTGAGCCCTTCACAAACCCCAACAATCTGGGACG	2791
Db	2650	CCCCCGCCCTTTGGCAGTAACTCTTCAGAGCCCTTCACAGACCCCTTAATTAATCTGGGCG	2709
QY	2792	ACTTCAGTTCACAGTGGCCCAAGAGCTCAAAAGAGAGCAGAAAACGGAACAATGGGATG	2851
Db	2710	ACTTCAGTTCACAGTGGCCCAACAGACAGATAGAAAACAGAAAGGATTAATCTGGAGTG	2768
QY	2852	CGTGGCGGCTCAACCTTCTCTGACCGTACCTTAGTCTGGCAGTTACGCGCAGAGATCAG	2911
Db	2770	CATGGGAGCCCAACCTCTCTCTCACCGTTCAAGTGCAGGCAAGTTAAAGCAGAGTCCG	2829
QY	2912	CCTTTACCCCAAGCCACAGCACTGAGCTCTCTCCCAATCTCCCGTCTGGGCAAGGCTGA	2969

[illegible]

Db 170 GAGCGCTGCAATTAGCAGAGTAAAGTAACAGAACCATAGGCTCAGTTTCCAAACCTTTTG 229
 QY 284 GTGGTACCTGGAGTGTCCGGGCGCTAACTGTGAGAGAAAGGCGCAAGCATGACAGAGT 343
 Db 230 GTGGCAGCTGGATATCTGGGCGCTAACTGTAGAGAAAGGCGCAAGCATGATCAGAGT 289
 QY 344 TCCTTAGCCTGAAAGCCGATAGCGGATTTTATTACTGTGTATCAAGCGAGAACTTTTTT 403
 Db 290 TCCATAGTTAAAGCCAAATATCTGGATTTCAATTACTGTGTATCAAGCTAAGAACTTTTTT 349
 QY 404 TCCATCTGGTTTACCTCAAGCTTCTTGAACAATAATATGGCGCTTACCGGATGAATA 463
 Db 350 TTCAATCTGGTTTACCTCAAGCTTCTTGAACAATAATATGGCGCTTACCGGATGAATA 409
 QY 464 ACGATGAGAGATGATCAAGTGAATTTTTCATAGCCATGAAAGCTTATCAAACTGAAGC 523
 Db 410 ATGATGAGAGATGATCAAGTGAATTTTTCATAGCCATGAAAGCTTATCAAACTGAAGC 469
 QY 524 TACAAGATATCAGCTCCCTCCACACTTCCCTGTATGAAACAGAAACAGTGGCTA 583
 Db 470 TACAAGATATCAGCTCCCTCCACACTTCCCTGTATGAAACAGAAACAGTGGCTA 529
 QY 584 TTTCCAGGACACGACATTTGGTATAGAGGATTTGCTAGATGCCACACTCAGAGCTG 643
 Db 530 TTTCTAGGACACGACATTTGGTATAGAGGATTTGCTAGATGCCACACTCAGAGCTG 589
 QY 644 TTGCTCTGTGCAATGGGCTTCATTCAGTTGTGGAATGTCTCAACCTTATGATCTT 703
 Db 590 TTGCTCTGTGCAATGGGCTTCATTCAGTTGTGGAATGTCTCAACCTTATGATCTT 649
 QY 704 CTGTGCTCTGAGGAGAGTGGCTCCCTGGCTAACGGGCGCTCCCTGATACAGCTC 763
 Db 650 CTGTGCTCTGAGGAGAGTGGCTCCCTGGCTAACGGGCGCTCCCTGATACAGCTC 709
 QY 764 TGCGTGGTTGGCCATCTCTGACGCCACATGGCCAAAGATTTCTTCCAGAGATCTG 823
 Db 710 TGCGTGGTTGGCTCTCTGACGCCACATGGCCAAAGATTTCTTCCAGAGATCTG 769
 QY 824 GTCCAGGCTCAATTAACAATACTAAGTTTACAGAAAGCAATATCTTGTATGCGCAGCG 883
 Db 770 GTCCAGGCTCAATTAACAATACTAAGTTTACAGAAAGCAATATCTTGTATGCGCAGCG 829
 QY 884 CCGCTCAGAGCAGAAATGGGCTGTGCTCAGTCATCAGGCTCAAAATACAGGAGTAT 943
 Db 830 TCCACCAAGTGGAGAGTGGGCTGTGCTCAGTCATCAGGCTCAAAATACAGGAGTAT 889
 QY 944 TCAAACAGCAGACAAAATACTAAGTGGACAATTACAGGTTCCCGCAGGCAAGAACTATT 1003
 Db 890 TCAATAGTCATGACAAAATACTAAGTGGACAATTACAGGTTCCCGCAGGCAAGAACTATT 949
 QY 1004 TCAATAGTCATGACAAAATACTAAGTGGACAATTACAGGTTCCCGCAGGCAAGAACTATT 1063
 Db 950 TTAATGAGTCAAGTTTACCAAGGCTCAGCTGCTTCAATATGGAATCTTTCAGCAATTG 1009
 QY 1064 ATCAAGATGAAAACTCACTGACAGAAATTTATCTTATGATGACCTATGATGTTG 1123
 Db 1010 ATCAAGATGAAAACTCACTGACAGAAATTTATCTTATGATGACCTATGATGTTG 1069
 QY 1124 CCAATGTCTGTGACGACCTGCGCGCTGTGCTCCAGAAATACATCCCTTCTTCTCA 1183
 Db 1070 CTATGTCTGTGCGCAACCACTGCGCTGTGCTCCAGAAATACATCCCTTCTTCTCA 1129
 QY 1184 GAAAGATTCGCTCCGGGAGTGGAGTGTCCGTATAGTCTTCTTCTGTGGATCAGAGG 1243
 Db 1130 GAAAGATTCGCTCCGGGAGTGGAGTGTCTCATATAGTCTCAACCTCTGTAGTCAAGG 1189
 QY 1244 TGCTGAGAGACCGTCTGACAGAGATGACAGAGC---CAGAGAGAAACTGCTTGTGA 1300
 Db 1190 TACCAAGAGAACCAATTTAGAGATGAAACAACAATTTGAAAAAATTTACTCTTAA 1249
 QY 1301 CATTTGAAGATTAAGAGGGGAGAACTTCGAGCGAGGACAGTGTGAGCTGAGAGAGGCC 1360

Db 1250 CATTGAAAGATTAAGAGCGGAGAACTTTGAACTGGCAACTGGAACTGGAGAAACGAA 1309
 QY 1361 GCCAAGCGCTCTTGGAGCAGCAGCGCAAAAGACGAGAGCGGTTGGCTCAGCTGAGCGC 1420
 Db 1310 GGCAGAGCTCTCTGGAGACAGCAGCGCAAGAGGAGAGCGGCTTGCCAGCTGGAGCGGG 1369
 QY 1421 CCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGGCCCAACGGCAGCTGAGC 1480
 Db 1370 CCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGGCCCAACGGCAGCTGAGC 1429
 QY 1481 TGGAGACAGCTGGAGAGAGCAGCGGAGCTGAGACCGCAGCGAGAGAGAGAGAGAG 1540
 Db 1430 TGGAGACAGCTGGAGAGAGCAGCGGAGCTGAGACCGCAGCGAGAGAGAGAGAGAG 1489
 QY 1541 AGAGATCGAAGAGCGCGAGCGCCGCAAAACGGGAACTGGAAAAGCAGCAGCACTTGAAT 1600
 Db 1490 AAGAAATTTAGAGGCGAGAGGCTGCAAAAACGGGAACTTTGAAAAGCAACGACACTTGA 1549
 QY 1601 GGGAGCGGAGCCGAGACAGAGAACTCTGATTCAGAGGAACAAGAGCAGAGGCGCACCG 1660
 Db 1550 GGGAGCGGAGCCGAGACAGAGAACTCTGATTCAGAGGAACAAGAGCAGAGGCACTAG 1609
 QY 1661 TGGTCTGGAAGCAAGAGAGAACTCTGAGTTTGAAGTTTGAAGCTCTGATGACAAA 1720
 Db 1610 TTGTCTGAAAGCAAGAGAAAGAAAGCTTTGAAATTTGAATTTGAAGCTCTTAAATGATAAA 1669
 QY 1721 AGCATCAGCTAGAGAGAAACCTTCAGAGATATCAGAGTGTCCATGSCAACCCAGAGGA 1780
 Db 1670 AGCATCAACTGAAAGAGAACTTCAGAGATATCAGAGTGTCCATGSCAACCCAGAGGA 1729
 QY 1781 AATTTGAGAGACGAAACAAGCTTAGAGAGTAAAGATTTGTAATGCAATCACCCTTACAGC 1840
 Db 1730 AATTTGAGAGACGAAACAAGCTTAGAGAGTAAAGATTTGTAATGCAATCACCCTTACAGC 1789
 QY 1841 AGCAGTTGAGAGATCTCAGCAATGCTTGAAGACTTTATTCAGAGAAACATATCTCA 1900
 Db 1790 AACAAATTCAGAGATCTCAGCAATGCTTGAAGACTTTATTCAGAGAAACATATCTCA 1849
 QY 1901 GTGACAGTTAAACAATCTCAGAGAAACAGTTTGCATAGAGACTGCTTCTTACCTCA 1960
 Db 1850 ATGACCAATTAACAATCTCAGAGAAACAGTTTGCATAGAGACTGCTTCTTACCTCA 1909
 QY 1961 AAGAGCCTTGGAAAGCAAGAGAGTGGCCCGCAGCAGACTCCCGGAGAGAGCTGAGCAGG 2020
 Db 1910 AAGAGCCTTGGAAAGCAAGAGAGTGGCCCGCAGCAGACTCCCGGAGAGAGAGCTGAGCAGG 1969
 QY 2021 TGGAGAGAGAGCCAGGTCAAAGCTGCGAGAGATTTGATTTTCAACAACAGCTGAAG 2080
 Db 1970 TGGAGAGAGAGCCAGGTCAAAGCTGCGAGAGATTTGATTTTCAATATATAGCTGAAG 2029
 QY 2081 AACTGAGAGAGATATAGCAAAACGCAACTCCAGAGCAGAGGTCCTGAGAGCAGCGC 2140
 Db 2030 AACTGAGAGAGATATAGCAAAACGCAACTCCAGAGCAGAGGTCCTGAGAGCAGCGC 2089
 QY 2141 GACTGAG 2182
 Db 2090 GACTGAG 2131

RESULT 12

AAZ34573 standard; cDNA; 2079 BP.

AAZ34573;

01-FEB-2000 (first entry)

Human SH3D1A cDNA clone 5.

DE SH3D1A gene; human; Down's syndrome; leukemia; cancer;
 XX megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;

lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 ss.
 Homo sapiens.
 Key Location/Qualifiers
 CDS 136..2079
 FT /tag= a
 MO9953062-A2.
 21-OCT-1999.
 16-APR-1999; 99WO-US08371.
 16-APR-1998; 98US-0082007.
 (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 Korenberg JR, Chen X;
 MPI, 1999-633829/54.
 P-PSDB; AA32157.
 Nucleic acid from the human SH3D1A gene and its products, useful for
 the diagnosis and treatment of myeloproliferative disorders and
 leukemia
 Claim 2: Fig 12; 99pp; English.
 This is the nucleotide sequence of a cDNA clone, termed clone 5,
 corresponding to a novel human SH3 gene, termed the SH3D1A gene,
 that contributes to the development of platelets and the
 pathogenesis of leukemias, both in general and in particular those
 involving the megakaryocytic lineage. The SH3D1A gene maps to the
 small candidate region for low platelets on chromosome 21.
 Sequencing of 5 different sizes of cDNA clone from foetal brain
 (see AA24570-74) suggests that at least 3 isoforms exist. The
 invention provides methods for the diagnosis and treatment of
 megakaryocytic abnormality, myeloproliferative disorder, platelet
 disorder, acute leukaemia, neural disorders, thrombocytopenia,
 platelet disorder on chromosome 21, low platelets in deletion for
 21, association of gains in chromosome 21 with leukaemias, neural
 abnormalities, dysfunctions and disorders including brain
 malformations and corresponding cognitive dysfunctions.
 microcephaly, lissencephaly, and colpocephaly. Methods are also
 provided for: suppressing cells unable to regulate themselves;
 screening for a somatic alteration in the SH3D1A gene; monitoring
 the progress and adequacy of a treatment; monitoring tumour risk
 progress or megakaryocytic abnormality; myeloproliferative disorder,
 haematopoietic disorder, platelet disorder or leukaemia; and
 treatment of a subject (including a prenatal subject) having
 megakaryocytic abnormality, myeloproliferative disorder,
 platelet disorder, leukaemia or neural disorder using a
 nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 other;
 Query Match 31.2%; Score 1584.4; DB 20; Length 2079;
 Best Local Similarity 85.8%; Pred. No. 0;
 Matches 1783; Conservative 0; Mismatches 291; Indels 4; Gaps 2;

304 GCCATTAAGTGGAGAAAGGGCCAGCATGACGAGAGTTCTTACCTGAAACCGATA 363
 181 GCCATTAAGTGGAGAAAGGGCCAGCATGATGACGAGTTCTTAAAGCCATA 240
 364 GCGGATTTATTACTGGTGTGATCAAGCGAGAACTTTTTCACATCTGGGTATCTCAG 423
 241 TCTGATTCATTACTGGTGTGATCAAGCTAAGAACTTTTTCATCTGGGTATCTCAA 300
 424 CCGTCTTAGCAAAATATGAGGCGCTAGCGGACATGATAAGATGAGATGATCAA 483
 301 CCTGTTTATGACAGATATGAGGCTAGCTACATGATATGATGATAAGATGATCAA 360
 484 GTGAATTTTCCATAGCCATGAGCTTATCAAACTGAGCTACAGGATATCACTCC 543
 361 GTGAGATTTTCCATAGCTATGAACTTATCAAACTGAGCTACAGGATATCACTACC 420
 544 TCCACACTTCCCTGTCATGATAAGACAGACCACTGGCTATTTCCATGACCAAGCATTT 603
 421 TCTGCACTTCCCTGTCATGATAAGACAGACCACTGGCTATTTCCATGACCAAGCATTT 480
 604 GGTATGAGGAGATTTGTAGCATGCACTCAAGCTGCTGCTGAGCCAAAGGGC 663
 481 GGTATGAGGAGATTTGTAGCATGCACTCAAGCTGCTGCTGAGCCAAAGGGC 540
 664 TCCATTTCCATTTGTTGGAATGTCTCAACCTTATGATCTTCTGCTCCAGCAGCAGTG 723
 541 TCCATTTCCATTTGTTGGAATGTCTCAACCTTATGATCTTCTGCTCCAGCAGCAGTG 600
 724 CCTCCCTGGCTAAACGGGGCTCTCCGCTATACAGCTGCTGCTGCTGGCTATCTT 783
 601 CCCCCCTGGCTAAACGGGGCTCTCCGCTATACAGCTGCTGCTGCTGGCTATCTT 660
 784 GGAGCCATATGCGCAAAAGTTCTTCTGAGCATGCTGCTGCTGCTGCTGCTGCTGCT 843
 661 GGAGCCATATGCGCAAAAGTTCTTCTGAGCATGCTGCTGCTGCTGCTGCTGCTGCT 720
 844 ACTAAGTTACAGAAAGGCAATATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
 721 ACTAATTTACAAAGGCAATATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 904 GCTGTGCTCAGTATCAAGGCTGATATCAAGGATATCAAGGATATCAAGGATATCAAG 963
 781 GCTGTGCTCAGTATCAAGGCTGATATCAAGGATATCAAGGATATCAAGGATATCAAG 840
 964 ATGATGACACTTATCAAGGCTGATATCAAGGATATCAAGGATATCAAGGATATCAAG 1023
 841 ATGATGACACTTATCAAGGCTGATATCAAGGATATCAAGGATATCAAGGATATCAAG 900
 1024 CAGGCTCAGGCTGCTTCAATATGGAATCTTCTGATGATGATGATGATGATGATGAT 1083
 901 CAGGCTCAGGCTGCTTCAATATGGAATCTTCTGATGATGATGATGATGATGATGAT 960
 1084 GCAGAAATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
 961 GCAGAAATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 1144 CCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203
 1021 CCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 1204 GGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1263
 1081 GGTATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 1264 GAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1320
 1141 GAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1200
 1321 GAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1380
 1201 GAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1260

QY 306 CATAACTGGAGAGAAAGGCCAAGCATGACACAGAGTTCTTAGCTGAAGCCGATAGC 365
 Db 394 CATTAATCTAGAGAGAAAGAGCGAAGCATGATGACAGTTCCATAGTTTAAAGCCAAATATC 453
 QY 366 GGGATTATTACTGATGATCAGCGAGAACTTTTTCATCTGGGTACCTCAGCC 425
 Db 454 TGGATTATTAATGAGTCAAGCTAGAAACTTTTTCATCTGGGTACCTCAGCC 513
 QY 426 TGTCTTAGCAAAATATGGCGCTTAGCGACATGATTAACATGAGAGATGATCAAGT 485
 Db 514 TGTTTTAGCAGAGATATGGGCACTAGCTGACATGAATTAATGATGAGAAATGATCAAGT 573
 QY 486 GGAATTTTCCATAGCCATGAAAGCTTATCAAACTGAAGCTACAAAGATACAGTCCCTC 545
 Db 574 GGAGTTTCCATAGCTATGAAACTTATCAAACTGAAGCTACAAAGATACAGTCCCTC 633
 QY 546 CAGACTTCCCTGCTGATGAAAGACCAACAGGCTATTTCCAGTGCACAGCATTTGG 605
 Db 634 TGGACTTCCCTGCTGATGAAAGACCAACAGGCTATTTCCAGTGCACAGCATTTT 691
 QY 606 TATAGAGGAGTGTAGCATGCGACCACTACAGAGTGTGCTCTGTGCCAATGGGCTC 665
 Db 692 ----- 691
 QY 666 CATTCAGTTTGAATGTCTCCACCTTAGTATCTTCTCCTCAGCAGAGCTGCC 725
 Db 692 ----- GCAAGTGTGCC 702
 QY 726 TCCCTTGGCTTAAGGGGCTCTCCCTCATACAGCTTGGCTTGGCGATCTCTGC 785
 Db 703 CCCCCTGGCTTAAGGGGCTCTCCCTGTATACAACTGTGCTGCAATTTGCTATCTCAG 762
 QY 786 AGCCACATGGCCAAAGAGTCTTCTCTCAGCAGATCTGGTCCAGGCTCACAATTAACAC 845
 Db 763 AGCCACATGGCCAAAGAGTCTTCTCTTATGATGTGCTGCAAGGCTCACAACTTAACAC 822
 QY 846 TAAATTACAAAGAGCACAATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAAGATGGCC 905
 Db 823 TAAATTACAAAGAGCACAATCATTTGATGTGCGCAGGCTCCAGCAGTGGCAGAGTGGCC 882
 QY 906 TGTGCTCAGTATCAAGGCTGAATATCAGGCGATTTATCAAGCAGCAAGCAAACTAT 965
 Db 883 TGTTCCTCAGTATCAAGGCTGAATATCAGGCGATTTATCAAGCAGCAAGCAAACTAT 942
 QY 966 GAGTGGCACTTAACAGGCTCCCGAGGAGAACTATTTCTGATGATCAAGTTTACCCCA 1025
 Db 943 GAGTGGCACTTAACAGGCTCCCGAGGAGAACTATTTCTGATGATCAAGTTTACCCCA 1002
 QY 1026 GGTCTAGCTGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAGAACTCAGC 1085
 Db 1003 GGTCTAGCTGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAGAACTCAGC 1062
 QY 1086 AGAAGAAATTTATCTTAGCTATGACCTAATTAATGATGTTGCTGATCTGATGAGCACTGCC 1145
 Db 1063 AGAAGAAATTTATCTTAGCTATGACCTAATTAATGATGTTGCTGATCTGATGAGCACTGCC 1122
 QY 1146 GCGCGTCTGCTCCAGAAATACATCCCTCTCTCAGAAAGGTTGCTCCGCGAGTGG 1205
 Db 1123 ACCGTCTGCTCCAGAAATACATCCCTCTCTTATAGAAAGGTTGATCTGCGAGTGG 1182
 QY 1206 GATGTCGCTGATAGCTCTTCTTCTGATCAGAGGCTGCTGAGAGAGCGTCTGACAG 1265
 Db 1183 TATATCTGTATAGGCTCAACATCTGTATGATCAGAGGCTCAGAGAGAACCACTTTTAGA 1242
 QY 1266 GGATGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAAAGGGGA 1322
 Db 1243 AGATGAACAACAATTAAGAAAGAAATTAAGTGAAGATTAAGAAAGGGGA 1302
 QY 1323 GAACTTGGAGAGAGAGCTGATGAGAGAGCGCGCAAGCGCTCTTGGAGCAGCA 1382
 Db 1303 GAACTTGGAGAGAGCTGATGAGAGAGCGCGCAAGCGCTCTTGGAGAGCAGCA 1362
 QY 1383 GCGCAAGAGAGAGAGCGGTTGGCTGAGTGGAGCGCGGAGCAGAGAGAGAAAGAGCG 1442

Db 1363 GCGAAGAGAGAGAGCGCTGCGCCAGCTGAGCGGCGGAGCAGAGAGAGAGCG 1422
 QY 1443 GGAGCGCCAGAGAGAGAGCGCAAGCGCAGCTGAGCTGAGAGAGAGAGAGAGCA 1502
 Db 1423 TGAAGCGCAGAGAGAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1482
 QY 1503 GCGGAGAGCTGAGCA 1562
 Db 1483 GCGGAGAGCTGAGCA 1542
 QY 1563 CGCAAAACGGGAACTGAGCA 1622
 Db 1543 TGAAGAGCGGAGCA 1602
 QY 1623 ACTCCTGAACTCAGCA 1682
 Db 1603 ACTTCTAATATCAAGCA 1662
 QY 1683 GACTCTGAGAGTTTGAAGTTAGAGAGCTTGAATGACAAAAGCATCAGTAGAGAGAGAGCA 1742
 Db 1663 GACTTGGAGTTTGAAGTTAGAGAGCTTGAATGACAAAAGCATCAGTAGAGAGAGAGCA 1722
 QY 1743 TGAAGATATCAGATGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1802
 Db 1723 TGAAGATATCAGATGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1782
 QY 1803 TGAAGATATCAGATGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1862
 Db 1783 TGAAGATATCAGATGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1842
 QY 1863 AATGCTTGAAGAGTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1922
 Db 1843 AATGCTTGAAGAGTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1902
 QY 1923 GCGAAGACGTTTGAAGCA 1982
 Db 1903 GCGAAGACGTTTGAAGCA 1962
 QY 1983 GCTGCGCCGAGCA 2042
 Db 1963 ACTGAGCTGAGCA 2022
 QY 2043 GCTGAGCA 2102
 Db 2023 ACTGAGCA 2082
 QY 2103 ACAGCACTTCCAGCA 2151
 Db 2083 GCAACCACTCCAGCA 2131

RESULT 14
 AA234574
 ID AA234574 standard; cDNA: 3231 BP.
 XX
 AC AA234574;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A cDNA clone 9.
 XX
 KW SH3D1A gene; human; Down's syndrome; leukemia; cancer;
 KW megalocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FT CDS 2..1927

FT /*tag= a
XX MO9953062-A2.
XX 21-OCT-1999.
XX 16-APR-1999; 99MO-US08371.
XX 16-APR-1998; 98US-0082007.
XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX korenberg JR, Chen X;
XX MPI; 1999-633829/54.
XX P-PSDB; AAY32158.
XX
XX Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT Leukaemia
PS
PS Claim 2; Fig 14; 99p; English.

CC This is the nucleotide sequence of a non-full-length cDNA (clone
CC 9) corresponding to a novel human SH3 gene, termed the SH3D1A
CC gene, that contributes to the development of platelets and the
CC pathogenesis of leukaemias, both in general and in particular those
CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
CC small candidate region for low platelets on chromosome 21.
CC Sequencing of 5 different sizes of cDNA clone from foetal brain
CC (see A424570-74) suggests that at least 3 isoforms exist. The
CC invention provides methods for the diagnosis and treatment of
CC megakaryocytic abnormality, myeloproliferative disorder, platelet
CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
CC platelet disorder on chromosome 21, low platelets in deletion for
CC 21, association of gains in chromosome 21 with leukaemias, neural
CC abnormalities, dysfunctions and disorders including brain
CC malformations and corresponding cognitive dysfunctions,
CC microcephaly, lissencephaly, and colpocephaly. Methods are also
CC provided for: suppressing cells unable to regulate themselves;
CC screening for a somatic alteration in the SH3D1A gene; monitoring
CC the progress and adequacy of a treatment; monitoring tumor risk
CC progress or megakaryocytic abnormality, myeloproliferative disorder,
CC haemopoietic disorder, platelet disorder or leukaemia; and
CC treatment of a subject (including a prenatal subject) having
CC megakaryocytic abnormality, myeloproliferative disorder,
CC platelet disorder, leukaemia or neural disorder using a
CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.

SO Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 other;
Query Match 27.4%; Score 1395.4; DB 20; Length 3231;
Best Local Similarity 70.7%; Fred. No. 0;
Matches 2359; Conservative 0; Mismatches 506; Indels 470; Gaps 19;

QY 1764 GGCACCCCAAGGAGCAAGATTGAGACGACCAAGTCTAGAGAGTGAATTTCTGA 1823
DB 1 GACCAACCCCAAGGAGCAAGATTGAGACCAAGTCTAGAGAGTGAATTTCTGA 60
QY 1824 AATCACCCTTACAGCAGCAGATTGACGAAATCTCAGCAAAATGCTTGGAAAGCTTATTC 1883
DB 61 AATCACCCTTACAGCAGCAGATTGACGAAATCTCAGCAAAATGCTTGGAAAGCTTATTC 120
QY 1884 AGAGAAACATATCTCGTACCGCAGTTAAACAAATGACACAGCAAGCTTGGATAGAG 1943
DB 121 AGAGAAACATATCTCGTACCGCAGTTAAACAAATGACACAGCAAGCTTGGATAGAG 180
QY 1944 CTCGCTTCTTACCTCTCAAAAGAGCTTGAAGCAAAAGAGCTGACCGGAGCAGAGCTCCG 2003
DB 181 TTCACTTGTAACTTAAAGAGCTTGAAGCAAAAGAGCTGACCGGAGCAGAGCTCCG 240
QY 2004 GGAGCAGCTGAGCAGAGTGAAGAGAGCAGCAGCTCAAGCTGAGAGATTTATGTTT 2063

DB 241 AGACCACTGATGAGTGAAGGAGAAAGAACTAGATCAAACTACAGAGATTGATATTT 300
QY 2064 CAACACCCAGCTGAAGGAACTGAGAGATACATAGCAAAACAGCACTCCAGAGACAG 2123
DB 301 CATATATAGCTGAAGGAACTTGAAGAAATACATATAGCAACACTCCAGAGACAAA 360
QY 2124 GTCCCTGAGGACGCGGACTGAAGCAAGAAAGACAGAGAGAGAGAGCTTGAATTGA 2183
DB 361 GTCCATGAGGCTGAACGACTGAAGCAAGAAAGAAAGAAAGATCTAGATTTAGA 420
QY 2184 GAAGCAAAAGAGAAAGCTGACAGAGAGTTCAGGAAAGGAGCAAGCAATGGCTGGAGCA 2243
DB 421 AAGCAAAAGAGAAAGCTGACAGAGAGTTCAGGAAAGGAGCAAGCAATGGCTGGAGCA 480
QY 2244 TGTGACGAGAG--GAGCAGCAGCCCGGAAACCCACAGAGAGAGACAGACTGA 2300
DB 481 TGTGACGAGAGAGCAGAGATCAGAGACCAAGAAATCTCACAGAGAGAGAAACTGA 540
QY 2301 GAGGAGACAGTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2360
DB 541 AAGGAGAGAGAGTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 2361 CAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGCAGTAAAGTGGCCACCCAGGAC 2420
DB 601 CAGCTGGGTGGCTTTTCCATCCGATCAGAGAGCAGTAAAGTGGCCACCCAGGAC 660
QY 2421 CTGCTTCAACAGAGAAAGGCGCTTACCATTTCTGACAGAGAGAGTGTAAAGTGT 2480
DB 661 CTGCTTCAACAGAGAAAGGCGCTTACCATTTCTGACAGAGAGAGTGTAAAGTGT 720
QY 2481 ATATTACGAGAGGCTGTACCCCTTTGAAATCCAGAGAGCAGAGAGATTCACATCCAG 2540
DB 721 GTATTACGAGAGGCTGTACCCCTTTGAAATCCAGAGAGCAGAGAGATTCACATCCAG 780
QY 2541 AGAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2600
DB 781 AGAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
QY 2601 GCTGAAAGGAGAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATG 2660
DB 841 ATTTAAAGAGAGAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATG 900
QY 2661 GATTCCTCCTCAGACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2720
DB 901 GATTCCTCCTCAGACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 2721 TCTGCTGAGACCCCTGCTCTTTCAGAGTGAATCTTTCAGAGTGAATCTTTCAG 2780
DB 961 CTTCGCTGAGACCCCTGCTCTTTCAGAGTGAATCTTTCAGAGTGAATCTTTCAG 1020
QY 2781 CAACCTGGGAGAGCTTCAAGTTCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2840
DB 1021 TAACTGGGAGAGCTTCAAGTTCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 2841 CAACCTGGGAGAGCTTCAAGTTCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2900
DB 1081 TAACTGGGAGAGCTTCAAGTTCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 2901 GAGAGATCAGCTTTTACCCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 2960
DB 1141 GAGAGATCAGCTTTTACCCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 1200
QY 2961 CCAGAGTGAAG 3020
DB 1201 CCAGAGTGAAG 1260
QY 3021 AGACAACTTAATTTTAAACAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3080
DB 1261 AGACAACTTAATTTTAAACAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1320
QY 3081 GTGGGTGTTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 3140
DB 1321 GTGGGTGTTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1380

QY 3141 CATTTCCAGGCGCCGTAAAGAAATCCACAGATCGATACCTGCGCTTACTGAAAGTCTGCG 3200
 DB 1381 CATTTCCAGGCGCCGTAAAGAAATCTACAGATCGATACCTGCGCTTACTGAAAGTCTGCG 1440
 QY 3201 TAGTCTAAAGAGATGCTTCCCGGCGCCCAAGCCAGCATTTCCCGGAGAAAGATTTAT 3260
 DB 1441 TAGTCTAAAGAGATGCTTCCCGGCGCCCAAGCCAGCATTTCCCGGAGAAAGATTTAT 1492
 QY 3261 TGGCATGTACACTACGAGAGTTCTGAGCAAGAGATTTAACTTTACGTCGAGGAGGATGT 3320
 DB 1493 ----- 1492
 QY 3321 GATTGTGTTACCAAGAAAGATGTGTACTGTGACGAGGACGTTGGCGGACAGTTCGG 3380
 DB 1493 ----- 1492
 QY 3381 AGTCTTCCCTTCTAACTATGTAGGCTTAAAGATTACAGAGGCTCTGGAACTGCTGGAA 3440
 DB 1493 ----- 1492
 QY 3441 AACAGGAGTTTAAAGAAAAAAGCTGAATTCGCGAGTTATGCTTCTTACGCTGCTAC 3500
 DB 1493 -----GAAATTCGCGAGTTATGCTTCTTACGCGCCAC 1527
 QY 3501 TGTGTCGGAACAACCTGCGCTGCTGCTGCGAGCTGATTCGATCCGAAAAAGAACCC 3560
 DB 1528 CGGCGCCGAGCAGCTCAGCTGCGCTGCTGCTGCTGATTTGATCCGAAAAAGAACCC 1587
 QY 3561 AGGTGATGTGTGGAAGAGAACTGCAAGCTCGAGGAAAAAGCCGCAATAGGCTGTT 3620
 DB 1588 AGGTGATGTGTGGAAGAGAACTGCAAGCTCGAGGAAAAAGCCGCAATAGGCTGTT 1647
 QY 3621 TCCAGCAATTTATGTAACTTCTAAGCCCGGAGCAAGCAAAATACCCCACTAGCT 3680
 DB 1648 CCGAGTTAATTTATGTAACTTCTAAGCCCGGAGCAAGCAAAATACCTCACAAGAGCC 1707
 QY 3681 ACCCAAGCCGAGCTGCGAGCAGCAGTGTGCGAGTGTATCGGATGTACGATTTACCGC 3740
 DB 1708 ACTTAAGTCAACAGCATTTAGCGCAGTGTGCGAGTGTATCGGATGTACGATTTACCGC 1767
 QY 3741 CCAAGAACGATGAGAACTGAGCTTTCAGCAAAAGCCGCAATGATTAAGTCTTCAACAAGA 3800
 DB 1768 GCAGATGATGATGATGATGCTTCAACAAGGCGCAGATCATCAAGTCTTCAACAAGA 1827
 QY 3801 GGCACCGGAGCTGTGGAAGAGAGATGATGAGGCAAGTGTGGCTCTTCCATCCATTA 3860
 DB 1828 GGCACCGGAGCTGTGGAAGAGAGATGATGAGGCAAGTGTGGCTCTTCCATCCATTA 1887
 QY 3861 TGTAAAGCTGACACAGACATGAGACCCGAGCCAGCAATGATTAATGTTGTCATCCCG 3920
 DB 1888 TGTAAAGCTGACACAGACATGAGACCCGAGCCAGCAATGATTAATGTTGTCATCCCG 1947
 QY 3921 CCTCAGGCTTGAAGTCTTCAAAAGAGACCCACTATCCATATCACTGCCAGAGGAGTG 3980
 DB 1948 CCTCAGGCTTGAAGTCTTCAAAAGAGACCCACTATCCATATCACTGCCAGAGGAGTG 1962
 QY 3981 ATGGAGATGACAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 4040
 DB 1963 ----- 1962
 QY 4041 GAAGAACTCACTGACAGAGAGTTTACTCTATTTGACCTTATGATGATGATGATGAT 4100
 DB 1963 ----- 1962
 QY 4101 CTGATGATCTGCGTGCAGAGGCAAGAAATTGCAAGACTGCAAGGCTGCTGCTCT 4160
 DB 1963 -----GTCTCT 1967
 QY 4161 TTTGGGCTTTCTTCTAGTCACTGAGACTGAC-CGGCGCCGCTTCAACAGCGGCGCTTTCAA 4219
 DB 1968 TTTGGGCTTTCTTCTAGTCACTGAGACTGACCTTTCCCGCCTTTGCAAGAGTCTTTCAA 2027

QY 4220 TAGTTTAAAGATTATTTTAAATGTGTATTTTACCTTTTAAATAAATCTCATCAT 4279
 DB 2028 TAGTTTAAAGATTATTTTAAATGTGTATTTTACCTTTTAAATAAATCTCATCAT 2087
 QY 4280 ACTTCTTGCTATTTTGGTTTAAACAAAACACCACTATCAAGAGTGTGCTGCGG 4339
 DB 2088 GACTTCTTTGCTATTTTGGTTTAAACAAAACACCACTATCAAGAGTGTGCTGCGG 2142
 QY 4340 ACGATTAAATGCTGTTCCGGGCGTACCGTTAACTGAGAGCTTGTCTGATCTTGCCT 4399
 DB 2143 GTGCTATTTAAATTTGTTCCAAATGTCCATTAATCGAGACTTGATGATTTTTCATTT 2202
 QY 4400 TGTCCAGTGTTCACCAACATATGTGTATTTGGGCTGTTCCTCG-CCGTAAGCA 4455
 DB 2203 TGTCCAGTGTTCACCAACATATGTGTATTTGGGCTGTTCCTCGCTTACATAGAAGTG 2262
 QY 4456 CAGAGAGATGGGTGATCCTGTTTGAATGTGTATGTAGACTGAGCCGATATGAA 4515
 DB 2263 CAGAGAGATGGGTGATCCTGTTTGAATGTGTATGTAGACTGAGCCGATATGAA 2321
 QY 4516 GGGGTATGCTTGTCTGTGACATCACTGTAC- TGTCCGAGATGTACATCTGACCG 4574
 DB 2322 GTGTTGTGCTTGTGTGTGTATCACTGTACCTGTGTGAGCATGTATATATCTGTA 2381
 QY 4575 AAGAATGACT-CTTCTCCATGCTTAAACCCACACCTGTACAGTCTCATCTACT 4633
 DB 2382 CATTAAGAAATGATTTCTTCCATGTGCAAACTTATACCTTGTGAGTGTATATAT 2441
 QY 4634 --GCATCATTTTACTTGTGACAGTACCTTGTAGCCAGCTGAGGAAAGCAC-CCATGTT 4690
 DB 2442 TGCAATTTATTTTATTTTGTGACAGTACCTTGTAGCCAGCTGAGGAAAGCTGTGTT 2501
 QY 4691 CCGTTGTGCTCAGATGTACCTGATTTGCGCGTGTGTTTATTTTCAATCTGCG 4750
 DB 2502 TGTGTCGCTCAGATTTATCTGTGTGATGTTGTTGTTGTTGTTTATTTTATTTGCG 2561
 QY 4751 ATGTCTTACACCACTAAATCTGTAGAGCCACCTGCGCGGCTGTATCATCTAGTA 4810
 DB 2562 GTGTTGTGATGATTAATTAATCTGTAGAGCCAC-ACCAGTGAAGTGTGTATCATCATTA 2620
 QY 4811 CCGACCGCT- -TTAGTCTGTGTAGCGTAAG- -TTATTTCCAGTGTCTTTTATGAA 4864
 DB 2621 TCACAGTCTCTTTTGT 2680
 QY 4865 - -TATCTGAAACAATATCTTCTTGAACAAGAAATGTATAGAGTCTCTGCA 4919
 DB 2681 TGACCTATTTGAAACAATATCTTCTTGAACAAGAAATGTATAGAGTCTCTGCA 2740
 QY 4920 ATTATTTCCAGTGTGTATCTTTTATCTAGACTGTGGGCTGTCTACATTAATAT 4979
 DB 2741 ATTATTT-CCATGTTTACATTTTAACTAGACTGT- -GGAATTTTACAGATTAATAT 2988
 QY 4980 GAAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5037
 DB 2799 GAAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2857
 QY 5038 -CCTTGAATATTAGTTGAAAGTCGGAAGAGAAAT 5071
 DB 2858 CTTTAAACACTAGTTGGAAGCTCTCAATAAAT 2892
 RESULT 15
 AAS84762
 ID AAS84762 standard; cDNA; 2874 BP.
 XX
 AC AAS84762;
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #20566.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 Food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.
OS
XX MO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR P-PSDB; ABG20575.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 20566; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 other;
SQ
Query Match 25.4%; Score 1291.4; DB 23; Length 2874;
Best Local Similarity 76.0%; Pred. No. 1e-303;
Matches 1778; Conservative 0; Mismatches 266; Indels 297; Gaps 3;
QY 1561 GCCGCAAAACGGGAATCGAAGGACGACCACTTGATGGAACGGAACCGGAGACAG 1620
DB 1 GCTGCAAAAACGGGAATCGAAGGACGACCACTTGATGGAACGGAACCGGAGACAG 60
QY 1621 GAACCTCTGTAATCAGAGAACAGAGAGAGAGGACCGCTGCTCTGAAAGGAAAGAG 1680
DB 61 GAACCTCTGTAATCAGAGAACAGAGAGAGAGGACCATTTTCTGAAAGGAAAGAG 120
QY 1681 AAGACTCTGAGTTTGTAGTGAAGAGCTCTGATGACAAAAGATCAGTACGAGAGAAA 1740
DB 121 AAGACTCTGAGTTTGTAGTGAAGAGCTCTGATGATGATGATGATGATGATGATGAT 180
QY 1741 CTTCAGAGATTCAGGTTGCACTGCGCAACCCAGAGGCAAGAAATTGAGAGCAGAACAG 1800
DB 181 CTTCAGAGATTCAGGTTGCACTGCGCAACCCAGAGGCAAGAAATTGAGAGCAGAACAA 240
QY 1801 TCTAGAGAGCTAAGATTTGCTGGAATCAGCCACTTACAGAGAGAGTTGCGAGAACTCAG 1860
DB 241 TCTAGAGAGCTTGAAGATTTGCGGAATCAGCCACTTACAGAGAGAGTTTACGAGATCTCAG 300

QY 1861 CAAATGCTTGGAGAGCTTATTCGAGAGAAAAGATATCTCAGTACGACAGTTAAAAAGATC 1920
DB 301 CAAATGCTTGGAGAGCTTATTCGAGAGAAAAGATATCTCAGTACGACAGTTAAAAAGATC 360
QY 1921 CAGCAGAAACAGTTTGCATAGAGAGCTGCTTTCTTACCTTCAAAAAGAGCTTTGAAAGCAAG 1980
DB 361 CAGCAGAAACAGTTTGCATAGAGAGCTTACCTTTCTTACCTTCAAAAAGAGCTTTGAAAGCAAG 420
QY 1981 GAGCTGCGCCCGGACAGAGCTCCCGGAGACAGTGGACGAGTGGAGAGAGACAGGTCAG 2040
DB 421 GAGCTGCGCCCGGACAGAGCTCCCGGAGACAGTGGACGAGTGGAGAGAGAGAGAGAGAG 480
QY 2041 AAGCTGACAGAGATTTGATGTTTTCACAAACAGCTGAGAGAACTGAGAGAGATACATAGC 2100
DB 481 AAGCTGACAGAGATTTGATGTTTTCACATATCAGCTGAGAGAACTGAGAGAGATACATAGC 540
QY 2101 AAACAGCACTCCAGAAAGCAGAGAGTCCCTTGGAGCGGAGCTTGAAGCAAGAAAGAGAG 2160
DB 541 AAACAGCACTCCAGAAAGCAGAGAGTCCCTTGGAGCGGAGCTTGAAGCAAGAAAGAGAG 600
QY 2161 GAGAGAGAGAGAGCTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
DB 601 GAGAGAGAGAGAGCTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 2221 AGGAGAGAGAGAGCTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2277
DB 661 AGGAGAGAGAGAGCTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 2278 CCGCAG 2337
DB 721 CCGCAG 780
QY 2338 AGAGCAAAACCGGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2397
DB 781 AGAGCAAAACCGGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 2398 GCTAAGCTGCGACACCGAG 2457
DB 841 GCTAAGCTGCGACACCGAG 900
QY 2458 GCACAG 2517
DB 901 GCACAG 960
QY 2518 CACGATGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAG 2577
DB 961 CACGATGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAG 1020
QY 2578 GAGCAG 2637
DB 1021 GAGCAG 1080
QY 2638 GAG 2697
DB 1081 GAG 1140
QY 2698 TCTGCGCCCTGCGCCCAAACTGAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2757
DB 1141 TCTGCGCCCTGCGCCCAAACTGAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 2758 TCTGAGCCCTGCGCCCAAACTGAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2817
DB 1201 TCTGAGCCCTGCGCCCAAACTGAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 2818 TCAACAG 2877
DB 1261 TCAACAG 1320
QY 2878 GTACCTAGTGTGCGCAGTTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2937
DB 1321 GTACCTAGTGTGCGCAGTTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

QY	286	GGTAGCCTGGATGTCGTGGGCCATATACGTGTGAGAGAAAGGCCAGATGACACAGAGTTC	34	445
Db	69	GGAGGGCCAAATATGTGGGGCTATTACATCTGGAGAAACGTACTAGATGATATTAACAGTTT	122	445
QY	346	CTTAGCCTGAGAGCCGATATGACGGGATTTATTAATCTGGTATCAAGCAGAGAACTTTTTC	405	445
Db	129	GATTAACCTTCAAACTTTCAGAGAGTTTATACATAACAGGATCAAGCCCGTACTTTTTCCTA	188	445
QY	406	CATTCCGGTTACCTCAGCCTGTCTTACGCAAAATATGGGCGCTAGCGAACAAGTAATAC	465	445
Db	189	CAGTCAAGTCTCCCGGCCCCGGTTTTTACCTGAAATATGGGCTTATCAGATGTGAACAG	244	445
QY	466	GATGGAAGATGATCAAGTGGAAATTTTCATATGACCATTAAGCTTAATCAACCTGAAGCTA	525	445

Db 249 GATGGAGATGAGCAGCAAGATTCTATAGCTATGAACTCATCAATTAAAGTTG 308
 QY 526 CAGAGATATAGTCCCTCCACACTTCCCTGTCATGAAACAGAACAGTGG---GCT 582
 Db 309 CAGGGCCCAAGAGCTGCTGATGCTCTCTCTATCATGAAACAAACCCCTATGTTCTCT 368
 QY 583 ATTTCCAGTGCACCAAGATTGGTATAGAGGAGATTGTCAGATG-----CCACCA 633
 Db 369 CCACTATCTCTGCTGTTTGGGATGGGAAGCATGCCCCAATCTGTCATTCATCAGCCA 428
 QY 634 CTACAGCTGTTGCTCTGTGGCCA-----TGGGCTCCATT 669
 Db 429 TTGCTCTCAAGTGCCTATAGCAACACCTTCTTCTGCTACTTCAGGAGCAGTAT 488
 QY 670 CCAATTGTTGAAATGTTCCACCTTATCTTCTGCTCCCTCAGAGAGAGCTCC 729
 Db 489 CTCCTCTTATGATGCTGCTCTCTCTAGTGCCTTCTGTTAGTACATCTCATACCAAA 548
 QY 730 CTGGCTAACGGGGCTCTCCCTGTCATACAGCTGCTGCTGCTTGGGATCCTGACGC 789
 Db 549 GGAATGCTCAGTCTCATTCAGCTTTATTCATTCCTTATCTTCAACATTCCTCAT 608
 QY 790 ACATGGCCAAAGAGTTTCTTCTCAGCAATCTGTCAGAGTCAACATTTAAACATAG 849
 Db 609 GCATCATCTTACAGCTGATGATGAGAGATTTGTTGTTGCTAGTATCCAGAAAGCCGAG 668
 QY 850 TTACAGAAAGCAACATTCATTCGATGTCGAGCGCCCTCCAG-----892
 Db 669 TCTCTGATGATTTAGATCTAGTAGCTCAACTTCTCACTGCTTCCCTCAGGGAAC 728
 QY 893 -----CAGAGAAATGGGCTGCTGCTCAGTCAATCAAGGCTGAAATCAGG 936
 Db 729 TCACCTAAGACAGGAGCTTGAAGTGGGAGTTCTTCAAGCTTCAAGATTTAAAGTACG 788
 QY 937 CAGTTATTTCAACAGCCAGCAAAACTATGATGAGCACTTAAACAGTCCCGCAGGAAG 996
 Db 789 CAAAAATTTTAATAGTCTAGCAAAAGGATGAGGAGTACCTCTCAGGTTTCAAGCTAG 848
 QY 997 ACTATTTCTCATGATCAATGATTTAACCCAGGCTGAGCTGCTTCAATATGGAATCTTTCT 1056
 Db 849 AATGCTCTTCTCAGTCAAAATCTCTCAAACTCAGCTACTATTTTGGACTGCTGCT 908
 QY 1057 GACATTTGATCAAGATGAGAACTCAGTCAAGAAATTTTCTAGTATGACCTAAT 1116
 Db 909 GACATCGATGTCGAGCACTTAAAGCTGAAGATTTTCTGCGGATGCACTCACT 968
 QY 1117 GATGTTCCATGCTGTCAGCACTGCTCCGCTGCTGCTCCAGAAATCATCTCTCT 1176
 Db 969 GACATGGCCAAAGCTGAGCAGCCACTACACTGACTGCTCCGAGCTTGTCTCCCA 1028
 QY 1177 TCTTTCAGAAAGTTCGCTCCGAGTGGGATGTCCTCATTAAGCTTCTTCTGAGAT 1236
 Db 1029 TCTTTCAGAGG-----GGAAAGCAAGTTGATTCGTTAAT 1064
 QY 1237 CAGAGCTGCTGAGAGCCGCTGTCAGAGATGAGCAGACAGCAAGAAACTGCTCT 1296
 Db 1065 GGAATCTGCTGCTTATATCAAAAAACCAAGAG---AAGAGCTCAGAAAGAACTGCA 1121
 QY 1297 GTGCAATTTGAAGTAAAGACGGGAGAACTTTCAGAGGAGGAGTGTGAGCTGGAGAG 1356
 Db 1122 GTTACTTTTGGAGCAAAACGAAAGCCAACTATATGAAAGAGAAACATGAGCTGGAGAG 1181
 QY 1357 CGCGCCCAAGGCTCTTGGAGCAGCAGAGGCAAAAGCAGGAGCGGTTGGCTCAGCTGAG 1416
 Db 1182 CGACGCCAAGTGTGATGAGCAGCAGAGGAGGCTGAAAGCCCAAGCCCAAGAGAG 1241
 QY 1417 CGCGCCGAGCAGAGAGAGAAAGACGGGAGCCAGAGCAGAGGCGCAAGCGCAGCTG 1476
 Db 1242 AAGAGAGTGGAGCGGAAACAGAGAGAACTGCAAGAGCAAGATGAAAGAGCAGCTG 1301
 QY 1477 GAGCTGGAAGCAGCTGGAAGAGCAGGAGGAGCTGAGCGCGCAGAGAGAGAGAG 1536
 Db 1302 GAGTTGGAGAAAGCTTGGAGAAACAGAGAGAGCTGGAGAGACAGCGGAGAGAGAGG 1361

QY 1537 AGGAGAGATCGAGAGCGGAGCGGCAAAACGGAACTGGAAGGCAAGCAACTT 1596
 Db 1362 AGAAAGGATATGAAAGACGAGAGGCGAGCAAAACAGGAGCTTATAGAGACAAACCGCTTTA 1421
 QY 1597 GAATGGAAACGGAACCGGAGACGGAACCTCTGATTCAGAGGAACAAAGGAGGAGGAGG 1656
 Db 1422 GAATGGAAAGACTCTGCTGGCAGAGAGCTCTCATGTAGAAACCGGGAACAAAGAGAC 1481
 QY 1657 ACCGTGCTCTGAGAGGAGGAGAAAGCTCTGAGTTTGAATGAGTTAGAGCTGGAATGAC 1716
 Db 1482 ATTTGACAGCTGAGCTCCAGAAAGAAAGTCTCCAGCTGGAATGGAAGCAGATGATGA 1541
 QY 1717 AAAAAATCTAGCTGAGAAAGGAAACCTTCAGATATCAAGTGTGACTGCAACCCAGAG 1776
 Db 1542 AAACATCAGCAGATCTCAGGAGACTCAAGATGCTCAAAATCAGAAAGCAAAAG 1601
 QY 1777 CAGAAATTTAGAGCAGAACCACTTAGAGAGCTAAGAAATGCTGAAATCACCACCTTA 1836
 Db 1602 ACTGAGCTAGAAAGTTTGGATTAACAGTGTACCTGGAATTTATGAAATCAAAACACTT 1661
 QY 1837 CAGCAGAGTTGCAAGAACTTCAGCAAAATGCTTGAAGACTTATTCAGAGAAACAGATA 1896
 Db 1662 CAACAAGAGCTTAAGAAATATCAAAATAGCTTATCTAGTCCCTGAGAGAGAGCTTA 1721
 QY 1897 CTGAGTGCAGTTAAACAAAGTCCAGAGAACGTTTGCATAGAGACTGCTCTTACC 1956
 Db 1722 TTAAACGAAATTAATAAACATCAGCTCAGTACACACTGATTCAGGATCAGTTTA 1781
 QY 1957 CTCAAAAGAGCTTGAAGCAAGAGAGCTGCGCGAGAGAGCTCCGAGAGCAGCTGAC 2016
 Db 1782 CTTCATTAAGAGTATCATAGAAAGAAAGAAATATAGCCAAAGACTTAAAGAACTTAGAT 1841
 QY 2017 GAGTGGAGAGAGAGCAGAGTCAAGCTGAGAGAGATTAATGTTTCAACACCACTG 2076
 Db 1842 GCTCTTGAAGAAAGAACTGATCTTACCTCAGAAATGATTCATTTAACTCACTG 1901
 QY 2077 AAGGAATCGAGAGATCATAGCAAAACAGCACTCCAGAAAGCAGAGGCTCTGAGGCA 2136
 Db 1902 AAGGAATCGAGAGAGATTAATACAGAGATTAAGCTTGAACAACTTCAATTAATC 1961
 QY 2137 GCGCAGCTGAAGCAGAAAGCAGAGAGAGAGAGCTGAGTTAGAGAAACAAA 2192
 Db 1962 AAACGTGACAAATTTGAAGAAATGAAAGAAAGATTAGAGCAAAAAA 2017

RESULT 2
 US-09-879-957-193
 ; Sequence 193, Application US/09879957
 ; Patent No. US20020034755A1
 ; GENERAL INFORMATION:
 APPLICANT: SPARKS, Andrew B.
 HOFFMAN, No. US20020034755A1h
 KAY, Brian K.
 FOWLES, Dana M.
 MCCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 USING SAME
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957
 FILING DATE: 13-Jun-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/630,915
 FILING DATE: 03-Apr-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELETYPE: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 193:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2873 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 193:
 US-09-879-957-193

Query Match 6.6%; Score 333.2; DB 10; Length 2873;
 Best Local Similarity 61.0%; Pred. No. 3.3e-76;
 Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

QY 2867 CTTCTCTGACCGTACCTAGTGGCCAGTTAGGGCAGATCAGCCTTACCCCGACCA 2926
 DB 541 CTTTTCATAAAGCTTAAATATCATGCGAAGAAAATCAGCCTTACCTGAACTG 600
 QY 2927 CAGCACTGCTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2986
 DB 601 TGTCCCTGG---ATCGTATCACTTATCATGACAGGACAGAGGTAGAAACTTAA 657
 QY 2987 AAGCGCAAGCCCTGTATCTCTGGAGACCCAAAGCAACCACTTAAATTTTAA 3046
 DB 658 AAGCAGAGGCGCTTGTCTTCTGACCTGCAAGAAAGATTAACCACTTGAACCT 717
 QY 3047 GTGACGTCATCAGCGTCTGGAACAGAACATGTCGTTTGGAGAACTCAAGGTC 3106
 DB 718 ATGCACTTATTTACTGTTTGGAGCAGCAAGAAATGTTGTTGGAGGTCATGGAG 777
 QY 3107 AGAAGGTTGTTCCCAAGTCTTACGTTGAAGCTCAATTCAGGCGCTTAAAGGAA 3166
 DB 778 GAAGAGATGTTTCCCAATCTTATGTCAAGATCATTCCTGGAGTGAAGTAAACGG 837
 QY 3167 CAAGCATGATCTGCGCCCTACTGAAAGTCTGCTAGTCTTAAAGAGTGGCTTCC 3226
 DB 838 AA-----GAACAGAAAGCTTGTATGACGCTTAATAAGAACTAACCTCCG 885
 QY 3227 CCGCAAGCAGCCATTCGCGGAGAAAGTTTATGCACTGATACATACGAAGTTCTG 3286
 DB 886 CAGCTTATTAAGT-----TGGAGAAAGATATATGACATTTATCAATTTCAAGTGG 939
 QY 3287 AGCAAGAGATTTTAACTTACAGCAAGGAGATGATTTGTTGTTTCAAGAAAGATG 3346
 DB 940 AACCTGAGATTTTGAATTTTCAAGAAAGTGAAGAAATTTGTTGATCCAGAAAGAT 999
 QY 3347 ACTGTTGAGCGGAACGTTGGCGCAAGTCCGAGTCTTCCCTTCTAATATGAGGC 3406
 DB 1000 AGTGTGAGCAGGAAGATTTGAGATGAGATGGAATTTTTCATCAAACTATGTCAA 1059
 QY 3407 TTTAAGATTCAGAGGCTCTGGAAGCTCTGGAAGAAAGAGGATTTAAGAAAGAACT 3466
 DB 1060 CAAAGATCAAGAGATTTTGGAGTGTAGCAAGTCTGAGCATCAAAATTAAGAACT 1119
 QY 3467 AAATTTGCCAGTTATTTCTTCAAGCTCTACTGTTCCCGAAACATCACTGAGCTC 3526
 DB 1120 AGATTTGTCAGGTACTTACAGCATATGTTGCTTGTGTTTCTGAACCACTTAACT 1179
 QY 3527 CTGGGACAGTGAATTTGATCCGGAAAAAGAAACCCAGGTGATGTTGGAGAGAACT 3586

DB 1180 CAGGACCTTATATTAATTTTAAAGAAAAATPACAGGGGTGTGGCAAGAGATTAC 1239
 QY 3587 AAGCTCAGAGGAAAAAGCGCAGATAGGTTGTTTCCAGCAATATATGCTAACTTAA 3646
 DB 1240 AGCCAGAGGAAAAAGGAGAGAAAGATGTTTCTCCGACATGTTAACTTTTGG 1299
 QY 3647 GCCCGGAACAGCAAAATCAACCCCACTGAGCTAACCAAGCCGAGTGCAGCCAG 3706
 DB 1300 GTCCAAAGTAGTAAAGAGCCACACTGCTTTTCATC-----CTG 1338
 QY 3707 TGTGACAGTATGCGGATGTAAGATTTACACCGCCGAAGCATGAGAACTACCTTCA 3766
 DB 1339 TATGTCAAGTATGCTATGATATGATATGATGAGCAAAATTAAGAAAGATGACT 1398
 QY 3767 GCAAGGCGCAGATCATCAAGCTCTCAACAGAGAGACCCGAGTGTGAAAGAGAA 3826
 DB 1399 CCAAGGACCACTATTTATGTTATGAACAAAGATGATCTGATTTGTTGGCAGAG 1458
 QY 3827 TCAGTGGCAAGTTGGGCTTCTCCCATCAATTTATGTAAGCTGACCAAGATGAC 3886
 DB 1459 TCACGCGGAGTACGTGCTCTTCTCCCAACTAGTTAAGATGACAGACATCAGAT 1518
 QY 3887 CCAGCCAGCAATGAATCA 3904
 DB 1519 CAAGTCAACAGTGACCA 1536

RESULT 3

US-09-764-868-125
 ; Sequence 125; Application US/09764868
 ; Patent No. US2002016871A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT232
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1510
 ; SOFTWARE: Patentlin Ver. 2.0
 ; SEQ ID NO 125
 ; LENGTH: 4210
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-868-125

Query Match 6.5%; Score 331.4; DB 9; Length 4210;
 Best Local Similarity 61.0%; Pred. No. 1.2e-75;
 Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2867 CTTCTCTGACCGTACCTAGTGGCCAGTTAGGGCAGATCAGCCTTACCCCGACCA 2926
 DB 918 CTTTTCATAAAGCTTAAATATCATGCGAAGAAAATCAGCCTTACCTGAACTG 977
 QY 2927 CAGCACTGCTCTCTCCCATCTCCCTCTGCGCCAGGCTGAAGAGTGAAGGCTTAC 2986
 DB 978 TGTCCCTGG---ATCGTATCACTTATTCATGACAGGACAAAGTGTAAACCTTAA 1034
 QY 2987 AAGCGCAAGCCCTGTATCTCTGGAGACCAAAAGACACACTTAAATTTTAA 3046
 DB 1035 AAGCAGAGGCGCTTGTCTTGGAGCTGCAAGAAAGATTAACCACTTGAACCTTCA 1094
 QY 3047 GTGACGTCATCAGCGTCTGGAACAGCAAGCATGTTGTTTGGAGAACTTCAAGGTC 3106
 DB 1095 ATGACATTTATTTCTGTTGAGAGCAGCAAGAAATTTGTTGTTGGAGTGCATGAG 1154
 QY 3107 AGAAGGTTGTTTCCCAAGTCTTACGTAAGAACTATTTCAAGGCGCTTAAAGAAAT 3166
 DB 1155 GAAAGAGATGTTTCCCAATCTTATGTCAAGATCATTCCTGGAGTGAAGTAAACGG 1214
 QY 3167 CAAGCATGATACGCGCTTACTGAAAGTCTGCTAGTCTTAAAGAGTGGCTTCCCGG 3226

```

Db 1215 AA-----GAACGAGAGCTTTGATGAGCTGTAAATAGAAACCTACCTCGG 1262
QY 3227 CCGGCAAGCCAGCATTCGCCGAGAGAGTTTATTTGCGATGACACATACGAGAGTTCTG 3286
Db 1263 CAGCTTATCAGT-----GGAGAGAGATATATATGACATTTATCCATATCAAGTCTGG 1316
QY 3287 AGCAGAGAGATTACCTTTTCAGCAGAGGAGTGTATTTGGTTTCCAGAGAAAGATGGTG 3346
Db 1317 AACCTGAGATTGACTTTCACAGAGAGTGAAGAAATATTTGGTGACCCAGAAAGATGGAG 1376
QY 3347 ACTGCTGAGAGGAGAGTGGGCGACAGTCCGAGTCTTCCCTTCACTATGAGGCG 3406
Db 1377 AGTGTGACAGAGAGATTGGAGTGTGAGATGAAAGTGAATTTTCCATCAACTATGTCAAAC 1436
QY 3407 TTAAGATTGAGAGGCTCTGGAACCTGCTGGGAAACAGGAGATTGAGAAAAAACTTG 3466
Db 1437 CAAGAGATCAAGAGAGTTTGGAGTGTCTGACAAAGTGTGGAGCATCAAAATAAAAACCTG 1496
QY 3467 AAATTGCGCAGGTTATTGCTTCTACGCTGTACTGTGCTCCGAAACAACCTACCTGGCTC 3526
Db 1497 AGATTGCTCAGGTAACTTCAGCATATGTGTGCTTGTGTTGTAACAACCTTAGCCTTGAC 1556
QY 3527 CTGCGCAGCTGATTCGTATCCGAGAAAGAAACCCAGGTGATGTTGGAGAGAGAACTGC 3586
Db 1557 CAGACAGATTATATTTATTTTAAAGAAAAATACAAAGTGGGTGGTGGCAAGAGATTAC 1616
QY 3587 AAGCTCAGAGGAAAAAGCCGAGATAGGGTGTGTTCCAGCAATTTATGCTAACTTTCTAA 3646
Db 1617 AGGCGAAGGAAAAAGCGACAGAAAGATGTGTTCTGCGCAGTCAGTAACTTTTGG 1676
QY 3647 GCGCGGAAACAAGAAATACCCCACTGAGTACCAGCAAGCCGAGTGCAGCGCAGCAG 3706
Db 1677 GTCCAAAGTAGAGAAAGAGCCACACCTGCTTTTCATC-----CTG 1715
QY 3707 TGTGCGCAGGTATGGAGATGATGATTAACACCGCCCAAGAGATGAGAACTGACCTTCA 3766
Db 1716 TATGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1775
QY 3767 GCAAGGCGCAGATCATCAGCTCCTCAACAGAGAGACCCGAGTGTGTAAGAGAGAG 3826
Db 1776 CCAAGGAGCAACTATTAAATGTAAGAAACAAGATGATCCTGATGTTGGCAAGAGAGA 1835
QY 3827 TCAGTGGGCAAGTGGGCTCTTCCATTCATTAATGTAAGAGTGCACAGACATGAGACC 3886
Db 1836 TCAAGGGGTGACGTGCTCTTCTTCTCAAACTAGTAAAGATGACAGACATCAGATC 1895
QY 3887 CCAGCCAGCAATG 3899
Db 1896 CAAGTCAACAGTG 1908

```

RESULT 4

```

US-09-879-957-39
; Sequence 39, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:

```

```

APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLER, Dana M.
McCONNELL, Stephen J.

```

```

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

```

```

NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

```

```

COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-879-957-39
Query Match 5.1%; Score 258; DB 10; Length 747;
Best Local Similarity 63.9%; Pred. No. 5.4e-57;
Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;
QY 3247 GGAGAGAGTTTATTTGCGATGATACATACGAGTCTTGAGCAAGAGATTAACTTT 3306
Db 115 GGAGAGAGATATTTGCACTTATTCATATTTCAAGTGTGAACCTGAGATTAACTTT 174
QY 3307 CAGCAAGGAGATGATTTGTGTTACCAAGAAAGTGTGATGCTGTGAGCGGAAACG 3366
Db 175 ACAGAAAGTGAAGAAATTTGTGTGACCCAGAAAGATGAGAGTGTGACAGAAATTT 234
QY 3367 GCGGCAAGTCCGAGTCTTCCCTTCTACTATGTAGAGCTTAAAGATTGAGAGGCTCT 3426
Db 235 GGAGATGAGAGGAAATTTTCCATCAAACTATGTCAAAACCAAGAGATCAAGAGATTT 294
QY 3427 GGAAGTCTGGGAAACAGGAGTTTGAAGAAAAACCTGAATTTGCCAGGTTATTTGCT 3486
Db 295 GGAAGTCTGAGCAAGTCTGGAGCATCAATTAATAAAACCTGAGATTGCTCAGGTAACTT 354
QY 3487 TCCTAGCTGTACTGTGTCGGAACAACCTCACTGCTCTGAGGAGCTGATTTCTGATC 3546
Db 355 GCATATGTTGCTTCTGTTGTTGAACAATTAGCTTGACAGAGAGTTAAATTTATTT 414
QY 3547 CGGAAAAAGAACCCAGGTGATGTTGAGAGAGAGATGCAAGCTCGAGGAAAAAGGCC 3606
Db 415 CTAAAGAAAAATACCAAGTGGGTGGGCAAGAGAGTTACAGGCGCAAGAAAAACGA 474
QY 3607 CAGATAGGCTGTTTCCAGCAATATATGTCAAACTTCAAGCCCGGAACAGCAAAATC 3666
Db 475 CAGAAAGAGTGTCTTCTGCGCAGTCATTTAACTTTGGGTCCAAGCAGTGAAGAGGCC 534
QY 3667 ACCCAACTGAGCTACCCAGAGACCGCAGTGCAGCCAGCAGAGAGTGCAGAGTGGGATG 3726
Db 535 ACACCTGCTTTCACTCT-----GTATGTCAGGTGATTTGCTATG 573
QY 3727 TACGATTACAGCCCAAGAGATGAGCAACTAGCTTACAGAAAGCCAGATCATCAAC 3786
Db 574 TATGACTATGACGAAATAAAGATGAGTCAAGTTTCTCAAGGAGCAACTATTAAT 633
QY 3787 GTCTCAACAGAGAGACCCGAGTGTGTAAGAGAGAGTCACTGGGCAAGTTGGGCTC 3846
Db 634 GTTATGAACAAAGATGATCTGATTTGGTGGCAGAGAGAGATCAACGGGGTGAAGTCTGCTC 693

```


PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO: 26948
LENGTH: 286
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
OTHER INFORMATION: NT HIT: AF114488.1, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUE 1.00e-112
US-09-664-761-26948

Query Match 3.9%; Score 198.8; DB 10; Length 286;
Best Local Similarity 87.2%; Pred. No. 7.9e-42;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2557 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGC 2616
DB 1 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGCA 60
QY 2617 GGATGCTTCCCTGCAAACTATGAGAAAAGATCCAGAAAATAGAGTCCCACTCCAGCC 2676
DB 61 GGGTGGTTCCTGCAAACTATGAGAAAATCCAGAAAATAGAGTCCCGCTCCAGTG 120
QY 2677 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCAACTGGCTTGGCTGAGACCTT 2736
DB 121 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCAACTGGCTTGGCTGAGACCTT 180
QY 2737 GCTCCTTTGGCAGTGAACCTTCTTGAGCCCTCCAGAACCCCAAGCACTGGGAGAGCTTC 2796
DB 181 GCCCTTTGGCAGTGAACCTTCTTGAGCCCTCCAGAACCCCAAGCACTGGGAGAGCTTC 240
QY 2797 AGTTCACAGT 2806
DB 241 AGTTCACAGT 250
RESULT 7
US-09-664-761-30453
Sequence 30453, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30453
LENGTH: 297
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUE 1.00e-112
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
US-09-664-761-30453

Query Match 3.9%; Score 198.8; DB 10; Length 297;
Best Local Similarity 87.2%; Pred. No. 8.1e-42;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2557 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGC 2616
DB 1 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGCA 60
QY 2617 GGATGCTTCCCTGCAAACTATGAGAAAAGATCCAGAAAATAGAGTCCCACTCCAGCC 2676
DB 61 GGGTGGTTCCTGCAAACTATGAGAAAATCCAGAAAATAGAGTCCCGCTCCAGTG 120
QY 2677 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCAACTGGCTTGGCTGAGACCTT 2736
DB 121 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCAACTGGCTTGGCTGAGACCTT 180
QY 2737 GCTCCTTTGGCAGTGAACCTTCTTGAGCCCTCCAGAACCCCAAGCACTGGGAGAGCTTC 2796
DB 181 GCCCTTTGGCAGTGAACCTTCTTGAGCCCTCCAGAACCCCAAGCACTGGGAGAGCTTC 240
QY 2797 AGTTCACAGT 2806

DB 241 AGCTCCACCGT 250

```

RESULT 8
US-09-864-761-17146
; Sequence 17146, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17146
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000311.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUUE 3.50e-01

```

```

; OTHER INFORMATION: NF HIT: AP114487.1, EVALUUE 1.00e-122
; OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUUE 4.00e-62
US-09-864-761-17146
Query Match 3.9%; Score 198.6; DB 10; Length 301;
Best Local Similarity 82.3%; Pred. No. 9.2e-42;
Matches 228; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
;
; 1298 TGACATTTGAATAGAGCGGAGAACTTTCGAGCGAGCGAGCTGTGAGTGGAGAGC 1357
; 1 TAACGTTTGAAGATPAGAGCGGAGAACTTTGAACCTGGCAACCTGGAGAGTGGAGAAAC 60
;
; 1358 GCCGCCAAGCCCTCTTGGAGCAGCAGCGCAAGCAGCAGCGCTTGTCTCAGCTGAGC 1417
; 61 GAAGCAGAGCTCTCTCTGAAACAGCAGCGCAAGGAGCGAGCGCTTGTCCAGCTGAGC 120
;
; 1418 GCGCGAGCAGAGAGGAAAGAGCGGAGCGCCAGCAGCAGAGCGCCAGCGGAGCTGG 1477
; 121 GGGCGAGCAGAGAGAGAGAGCGGAGCGCTTGAAGCGCGAGCAAGCGCAAAAGCAACTGG 180
;
; 1478 AGCTGAGAGACAGCTGAGAGAGCGGAGCGGAGCTTGAAGCGCGAGCGAGAGAGAGAGA 1537
; 181 AACTGAGAGACCACTGGAAGAGCGGAGCGGAGCTTGAAGAGCGAGAGAGAGAGAGAGA 240
;
; 1538 GGAAGAGATGAGAGCGCGGAGCGCCGCAAAACGGCA 1574
; 241 GAAAGAAATTGAGAGCGCGAGGCTTGAAGCAGCGCGAGA 277
;
RESULT 9
US-09-864-761-17644
; Sequence 17644, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17644
LENGTH: 263
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUE 3.00e-62
OTHER INFORMATION: SWISSPROT HIT: O15601, EVALUE 2.80e-01
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 1.00e-122
US-09-674-761-17644

Query Match 3.9%; Score 197.4; DB 10; Length 263;
Best Local Similarity 84.4%; Pred. No. 1.7e-41;
Matches 222; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1298 TGAATTGTAAGATAAGAGCGGAGAACTTGCAGCGAGCACTGTGAGCTGGAGAAC 1357
DB 1 TAACTTTGAAGATAAGAGCGGAGAACTTGAACGTGCAACTGGAACTGGAGAAAC 60
QY 1358 GCGCGCAAGCGCTTGTGGAGCAGCAGCGCAAGAGAGAGCGGCTGCTCACTGGAGC 1417
DB 61 GAAGGCAAGCTTCTCTGGAACAGCAGGAGAGAGAGAGAGCGCTGCGCCAGCTGGAGC 120
QY 1418 GCGCGCAGCAGAGAGAGAGAGAGCGGAGCGCCAGAGAGAGAGCGCGAGCTGG 1477
DB 121 GCGCGAGCAGAGAGAGAGAGAGCGGAGCGCCAGAGAGAGAGCGCGCAAAAGCAACTGG 180
QY 1478 AGCTGGAAGAGCGCTGGAGAGAGCGGAGCTGGAGCGCGCAGCGAGAGAGAGAG 1537
DB 181 AACTGGAAGAGCAACTGGAGAAAGCAGCGGAGCTAGAAAGGAGAGAGAGAGAGAG 240
QY 1538 GGAAGAGATCGAGAGCGCGAG 1560
DB 241 GGAAGAAATTTGAGAGCGCGAG 263

RESULT 10
US-09-674-761-10314
Sequence 10314, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 10314
LENGTH: 480
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
US-09-674-761-10314

Query Match 3.8%; Score 192; DB 10; Length 480;
Best Local Similarity 81.6%; Pred. No. 6.5e-40;
Matches 222; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2523 TGAGATCAATCCAGCAGAGATATGATGATGATGAAGCCAGCTGGAGAGCC 2582
DB 209 TTAGTTCTACTTAAGCTGAGACTTTTCCACAGGTGATGAAGCCAACTGGAAGC 268
QY 2583 AGATGCTTGGAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2642
DB 269 CGGCTGCTTGGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 328
QY 2643 AAGATTCAGAAATAGGTTCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2702
DB 329 GAAATCCAGAAATAGGTTCCCGCTCCAGTGAAGAACAGTACGATTCACATCTGC 368
QY 2703 CCGTCCCCCAAACTGCTCTGCTGAGACCCCTGCTCTTCCAGTACCTCTTCTGA 2762
DB 389 CCGTCCCCCAAACTGCTCTGCTGAGACCCCGCTCTTCCAGTACCTCTTCTGA 448
QY 2763 GCGCTCCAGCAAGCCCAAGCACTGGGAGACT 2794
DB 449 GCGCTCCAGCAAGCCCTTAATATGAGTGGCGGACT 480

RESULT 11

US-09-864-761-333

Sequence 333, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1

CURRENT FILING DATE: 2001-05-23

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-09-27

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

Best Local Similarity 84.6%; Pred. No. 5,4e-37;
Matches 203; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 1297 GTGACATTGAAGTAAGAACCGGAGAACTTCGAGCGAGCGAGTGTGAGTGAAG 1356

Db 247 GTAACGTTGAAGTAAGAACCGGAGAACTTCGAGCGAGCGAGTGTGAGTGAAG 306

OY 1357 CGCCGCCAAGCGCTCTTGAGCAGCAGCGCAAGAGCAGGCGTTGCTCAGCTGGAG 1416

Db 307 CGAAGCAAGCTCTCTTGAAACAGCAGCGCAAGAGCAGGCGCTGCGCCAGCTGGAG 366

OY 1417 CGCCGCCAAGCGCGAGAGAAAGCGGAGCGCGCAGGAGCAGGAGCGGCGAGCTG 1476

Db 367 CGGCGGAGCGAGAGAGAGAGAGCGGAGCGCGCAGGAGCGGAGCGGAGCGGAGCTG 426

OY 1477 GAGCTGGAAGCAGCTGGAAGCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCTG 1536

Db 427 GAATGGAAGCAGCTGGAAGCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCTG 486

RESULT 12

US-09-864-761-864

Sequence 864, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1

CURRENT FILING DATE: 2001-05-23

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-09-27

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 864

Query Match 3.6%; Score 180.8; DB 10; Length 486;


```

: LENGTH: 487
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000049.1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
US-09-664-761-864

Query Match          3.6%; Score 180.8; DB 10; Length 487;
Best Local Similarity 84.6%; Pred. No. 5,4e-37;
Matches 203; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1297 GTGACATTTGAAGATTAAGACCGGAGAACTTGACGAGCAGTGTGAGCTGAGAG 1356
DB 248 GTACGTTGAATAGAAAGCGGAGAACTTGAACGTGCAACTGGAAGAGAGAA 307
QY 1357 CGCCGCAAGCGCTCTTGAGCAGCAGCGCAAGAGCAGAGCGGTTGCTCAGCTGAG 1416
DB 308 CGAAGCGAACTCTCTGGAACAGCAGCGCAAGAGCAGAGCGCTGAGCGAGCTGAG 367
QY 1417 CGGCGCAGCAGAGAGAGAAAGCGGAGCGCAGAGAGAGAGAGCGGCGAGCTG 1476
DB 368 CGGCGCAGCAGAGAGAGAGAGAGAGCGGAGCGCAGAGAGAGAGAGAGAGAGAG 427
QY 1477 GACCTGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
DB 428 GAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487

RESULT 13
US-09-864-761-311
: Sequence 311, Application US/09864761
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US 09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668

```

```

: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 311
LENGTH: 475
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000193.1
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-311

Query Match          3.2%; Score 163; DB 10; Length 475;
Best Local Similarity 80.9%; Pred. No. 2.3e-32;
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2523 TGACATCACCATCCAGCAGAGATATATGTCATGTGATGAAGAGCCAGCTGAGAGCC 2582
DB 241 TTAGTTCTACTTAAGGTGTGACTTTTCCACAGGTGATGAAGAGCCAACTGAGAGACC 300
QY 2583 AGATGCTTGAAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2642
DB 301 CGGCTGCTTGAAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 2643 AAAGATTCCAGAAATAGAGTTCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2702
DB 361 GAAATCCCAAGAAATAGAGTTCCGCTCCAGTGAACCAAGTACTGATTCAACATCTGC 420
QY 2703 CCCGCCCCCAAACTGAGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2757
DB 421 CCTGCCCCCAAACTGAGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475

RESULT 14
US-09-864-761-13884
: Sequence 13884, Application US/09864761
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US 09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456

```

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13884
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000117.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
US-09-864-761-13884

```

```

Query Match      3.2%; Score 163; DB 10; Length 475;
Best Local Similarity 80.9%; Pred. No. 2.3e-32;
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

```

QY 2523 TGAGATCATCATCCAGCCAGAGATATATGATCATGATGATGAAGCCAGAGAGCC 2582
DB 241 TTACTTTACTTAAGCTGTGACTTTTCCACAGGTGATGAAGCCAACTGAGAAC 300
QY 2583 AGGATGCTTGAAGAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2642
DB 301 CGGCTGCTTGAAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 2643 AAAGATTCAGAGAAATAGGTTCCCACTCCAGCCAAACAGTACGATGATCTGC 2702
DB 361 GAAATATCCAGAGAAATAGGTTCCCACTCCAGCCAAACAGTATCAATCTGC 420
QY 2703 CCCGCCCCCAAACTGCTGCTGCTGAGAGCCCTGCTCTTGGCAGTGAAGCTCT 2757
DB 421 CCTGCCCCCAAACTGCTGCTGCTGAGAGCCCTGCTCTTGGCAGTGAAGCTCT 475

```

RESULT 15

```

US-09-884-441-5
; Sequence 5, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-884-441-5

```

```

Query Match      3.0%; Score 151.4; DB 10; Length 531;
Best Local Similarity 61.3%; Pred. No. 2.6e-29;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

```

```

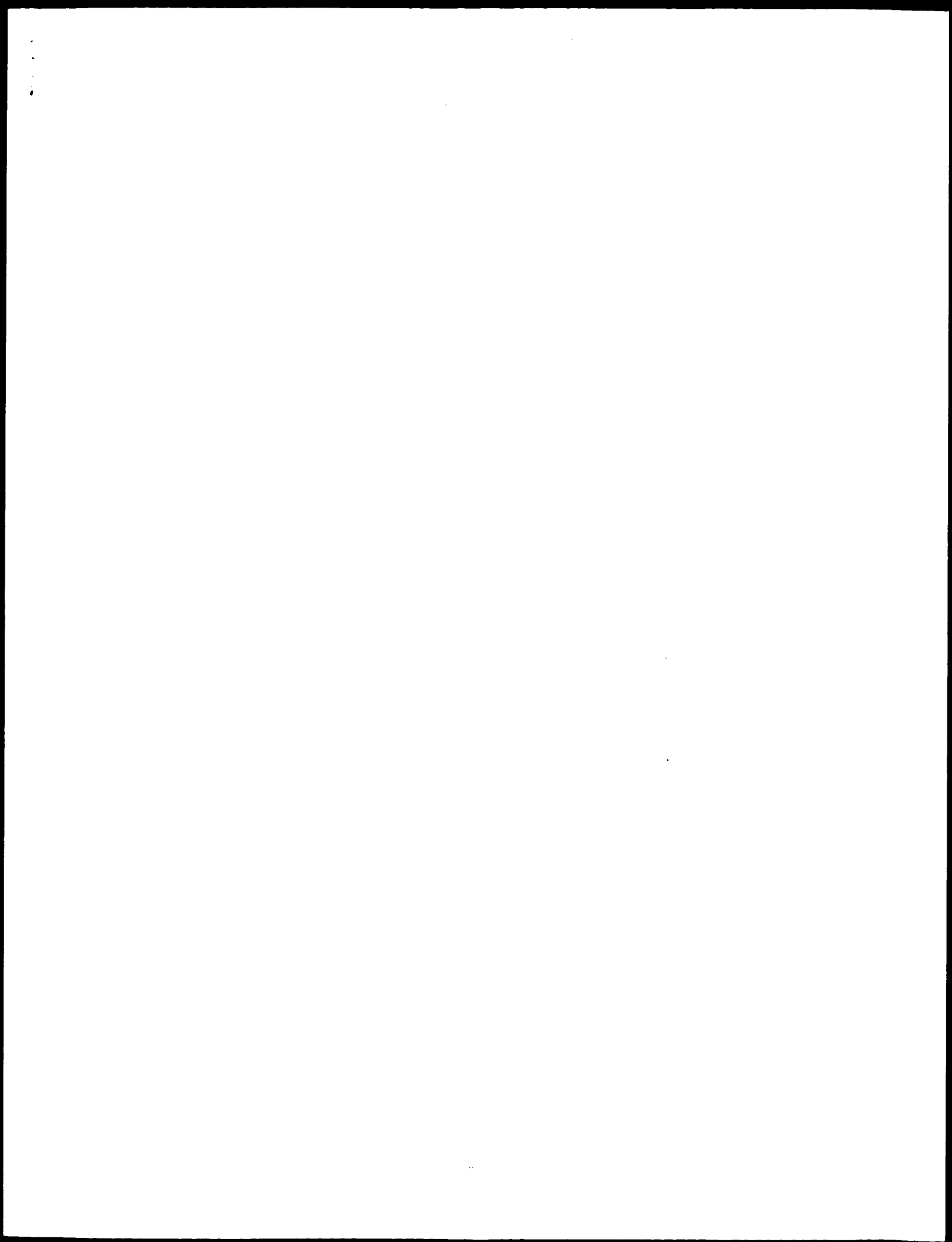
QY 226 GCGTCGATCAGCAAGTGAAGCTATAGAACCATGGCTGCTCCACACCTTGGT 285
DB 9 GCGTGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 68
QY 286 GGTACCTGAGATGCTGGGAGCATTAAGTGGAGAGAGAGAGAGAGAGAGAGAG 345
DB 69 GAGAGGCAAAATATGTGGCTATTAATCATCTGAAGAGAGAGAGAGAGAGAGAG 128
QY 346 CTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
DB 129 GATACCTCAAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
QY 406 CAATCTGGATTAAGTCAAGCTCTCTAGACAAATATGGGAGAGAGAGAGAGAG 465
DB 189 CAGTAGGCTGCGCGCCCGCTTTAGCTGAATATGGGAGAGAGAGAGAGAGAG 248
QY 466 GATGAGAGAGAGATCAAGTGAATTTCCATAGCCATGAAGCTTATCAAGTGA 525
DB 249 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
QY 526 CAAGATATCAAGCTCCCTCCCACTTCCCTCTCATGAAGAGAGAGAGAGAGAG 585
DB 309 CAGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
QY 586 TCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645
DB 363 TTCTCTCACTAATCTGCTGCTGTTGGAGATGGAAGAGAGAGAGAGAGAGAG 422
QY 646 GCTCTGTGCC 656
DB 423 CAGCATTGCC 433

```

```

Search completed: March 14, 2003, 11:36:48
Job time : 967.707 secs

```



Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Parvahu, Parvaneh Saedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.jnl.gov>
Series: IRAL Plate: 40 Row: n Column: 1
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796
This clone has the following problem: no cloning site / microdeletion.

FEATURES

Location/Qualifiers

1..2126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4899011"
/tissue="Pancreas, epithelioid carcinoma"
/clone_lib="NIH MGC 42"
/lab_host="DH10B-R"
/note="Vector: pDHB7"

BASE COUNT 712 a 462 c 517 g 435 t

ORIGIN

Query Match 31.4%; Score 1598.2; DB 11; Length 2126;
Best Local Similarity 85.3%; Pred. No. 0;

Matches 1822; Conservative 0; Mismatches 298; Indels 15; Gaps 3;

68 GCAGAGGAGATCCCGAGCGGGCTCCGGAGCGCGGAGGCGAGGCGGGCGG 127
Db 1 GCAGAGGAGATCCCGAGCGGGCTCCGGAGCGCGGAGGCGGGCGG 49
128 GATGATGATCGGCGCGGCGGAGCTCGGCTTCCTCCGCGCGGCGGCGGCTGAT 186
Db 50 GATGATGATCGGCGCGGCGGAGCTCGGCTTCCTCCGCGGCGGCGGCTGAT 109
187 TTGTGTAGGGGCGCGCGCGGCGGAGTGAAGCGCTGATGAGCAAGGTAA 246
Db 110 TTGTGTAGGGGCGCGCGGCGGAGTGAAGCGCTGATGAGCAAGGTAA 169
247 GATGATGAGCAAGTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 306
Db 170 GATGATGAGCAAGTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 229
307 ATAACTGTGAGGAGGAGGCGGAGGAGTGAAGCGGCTGATGAGCGGATAGG 366
Db 230 ATAACTGTGAGGAGGAGGCGGAGGAGTGAAGCGGCTGATGAGCGGATAGG 289
367 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 426
Db 290 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 349
427 GATGATGAGCAAGTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 486
Db 350 GATGATGAGCAAGTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 409
487 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 546
Db 410 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 469
547 AACCTGCGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 606
Db 470 AACCTGCGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 529
607 ATAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
Db 530 ATAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 589
667 ATTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 726
Db 590 ATTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 649

727 CCCCTGGCTAAGGGGCTCCCGCTCATAGAGCTGTGCTGCTTGGCATCTTCCA 786
Db 650 CCCCTGGCTAAGGGGCTCCCGCTCATAGAGCTGTGCTGCTTGGCATCTTCCA 709
787 GCCATGAGCCAAAGAGTCTTCTTCCAGCATCTGCTGAGGAGTCAAACTAACT 846
Db 710 GCCATGAGCCAAAGAGTCTTCTTCCAGCATCTGCTGAGGAGTCAAACTAACT 769
847 AAGTTACGAAAGGCAATCAATTGATGCGCAGCGCCCTCCAGCAGAGATGGGCT 906
Db 770 AAGTTACGAAAGGCAATCAATTGATGCGCAGCGCTCCAGCAGAGATGGGCT 829
907 GTCCTGATCATCAAGGCTGAAAATACAGGAGTATTTCAACCCAGCAAACTATG 966
Db 830 GTCCTGATCATCAAGGCTGAAAATACAGGAGTATTTCAACCCAGCAAACTATG 889
967 AGTGACACTTAACAGGCTCCAGCAGAACTATTTCTGACATGCAATCAAGTTACCCAG 1026
Db 890 AGTGACACTTAACAGGCTCCAGCAGAACTATTTCTGACATGCAATCAAGTTACCCAG 949
1027 GCTGAGCTGGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
Db 950 GCTGAGCTGGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1009
1087 GAAGAAATTTATCTAGCTATGCACTAATGATGATGATGATGATGATGATGATGAT 1146
Db 1010 GAAGAAATTTATCTAGCTATGCACTAATGATGATGATGATGATGATGATGATGAT 1069
1147 CCGCTGCTGCTCCAGAAATATATCTCTTCTTCAAGAAAGTTCGCTCGGAGTGG 1206
Db 1070 CCGCTGCTGCTCCAGAAATATATCTCTTCTTCAAGAAAGTTCGCTCGGAGTGG 1129
1207 ATGTCCTGATTAAGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGAT 1266
Db 1130 ATGTCCTGATTAAGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGAT 1189
1267 GATGAGCAGCAGC---CAGAGAGAACTGCTGTGACATTTGAAGTAAAGACGGGAG 1323
Db 1190 GATGAGCAGCAGC---CAGAGAGAACTGCTGTGACATTTGAAGTAAAGACGGGAG 1249
1324 AACTTCAGGAGGAGGAGTGTGAGCTGAGAGAGCGCCAGACGCTTTGAGCAGCAG 1383
Db 1250 AACTTCAGGAGGAGGAGTGTGAGCTGAGAGAGCGCCAGACGCTTTGAGCAGCAG 1309
1384 CGCAAAAGAGCAGAGCGGTTGCTCAGCTGAGCGCGCGAGCAGAGAGAGAGAGCGG 1443
Db 1310 CGCAAAAGAGCAGAGCGGTTGCTCAGCTGAGCGCGCGAGCAGAGAGAGAGAGCGG 1369
1444 GAGCGCAGAGCAGAGCGGCTCAAGCGGAGCTGAGCTGAGAGAGAGCTGAGAGAGCAG 1503
Db 1370 GAGCGCAGAGCAGAGCGGCTCAAGCGGAGCTGAGCTGAGAGAGAGCTGAGAGAGCAG 1429
1504 CGGAGCTGAGAGCGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563
Db 1430 CGGAGCTGAGAGCGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1489
1564 GCAAAAGGAGAACTGAGAAAGCAGCACTTGAATGAGAAAGAGAGAGAGAGAGAGAG 1623
Db 1490 GCAAAAGGAGAACTGAGAAAGCAGCACTTGAATGAGAAAGAGAGAGAGAGAGAGAG 1549
1624 CTCTGATATGAG 1683
Db 1550 CTCTGATATGAG 1609
1684 ACTCTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAG 1743
Db 1610 ACTCTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAG 1669
1744 CAGGATATCAGGTGTGACTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
Db 1670 CAGGATATCAGGTGTGACTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729

BASE COUNT	703 a	451 c	492 g	433 t	ORIGIN
					/clone_lib="NIH_MGC_68" /lab_host="VDH10B" /note="Vector: PCMV-SPORT6"
Query Match					31.1%; Score 1582; DB 11; Length 2079;
Best Local Similarity					85.8%; Pred. No. 0;
Matches 1780; Conservative					0; Mismatches 290; Indels 4; Gaps
QY	129	GATGCTGTGCGCGCTGCGGACTCGGCGCTTCTCGC	-CGGCGCTGCGCGCTGCACGTATT	187	
Db	1	GATGCTGTGCGGCGCTGCGGCTCTGCGCTCCACGCGCGCTAGCGGCACTGATT	60		
QY	188	TGCTGTAGAGGGCGCGCGCGCACCGCCCGAGATGAGGCGTGCATCAGCAAGTGTAAC	247		
Db	61	TGTCCTCGGAGCGCGCACGCGCGACCCCGAGATGAGGCGTGCATTAGCAAGTAAAA	120		
QY	248	GTAATAGAACCATGCGCTCAGTTTCCCAACCTTTGCGTGTAAGCTTGAGTCTCGGSCCA	307		
Db	121	GTAACAGAACCATGCGCTCAGTTTCCCAACCTTTTGGTGAGCAGCTGAGTATCTGGGCCA	180		
QY	308	TAACTGTGAGGAAGAAGGCCAAGCAAGTACACAGTCTTCAAGCTGTAAGCCGATACGG	367		
Db	181	TAACTGTAGAGGAAGAAGCAAGCAAGTACACAGTCTCAATAGTTTAAGCCATATCTG	240		
QY	368	GATTATTACTGTGTGATCAAGCGAGGAACCTTTTTTCCAATCTGGGTTACCTCAGCGCTG	427		

QY 428 TCTTAGCACAATAATATGGCGCGCTAGCCGACATGAATTAACGATGGAAGATGGATCAATGG 487

Db 301 TTTTAGCACAGATATGGGCACTAGCTGACATGAATAAATGATGGAAGAAATGGATCAATGG 360

QY 488 AATTTTCATAGCCATGAAAGCTTATAAATCTAAGCTCAAGGATATCAGTCCCTCCCA 547

Db 361 AGTTTCCATAGCTATGAACTTATATAAACTAAGCTCAAGGATATATAGTTACCCTCTG 420

QY 548 CACTTCCCCCGTCAATGAACACGACCAACGATGGCTATTTCCAGTGCACACAGATTTGGTA 607

[illegible]

Ov 728 CCCGAGCTAACGGGGCCTCCCGGTATACAGCGCMCTGCCTGCCGTTTGCGCATCTGCAG 787
Db 541 TTCGAGTIGTAGAAAGTCACCAACCCTAGAATCTCTGAATCCCACAACACACTGCCCC

Db	601	CCCTGGCTTAAAGGGGGCTCCCTCGTATATACACCTTGGCTTCATTTGGCTATCTCTGAG	660
Qy	788	CCACATGGCCAAAGATTCTTCCCTTCAGCAGATCTGGTCCAGGGTCAACAATTAAACACTA	847
Db	661	CCACATTCGCCAAAGATTCTTCCCTTTAGTAGATCTGGTCAGAGGTACACAACCTAAACACTA	720
Qy	848	AGTTCCACAAGGCACATTCATTCGATGTGGCCAGGCCCCCTCCACACAGACAGATGGGCTG	907
Db	721	ATTTCCTAAATGGCTCACTCATTTTATTTATTTGGTCCACAGTGTCCACACAGGGGATGGGCTG	780

Qy	908	TTCCCTCAGTTCATCAAGGCTGTAATACAGGCGAGTATTTCACAGCCACGCAAAACTATGA	967
Qy	781	TTCCCTCAGTTCATCAAGCTGAATACAGGCAATATTTCATGTGCTATGCACAAACTATGA	840
Qy	968	GTGCACTTCAACAGGTCCCGAGCAAGACTATTCTCATGCAATCAAGTTACCCGAG	1027

Db 841 GTGGACCTTAACAGGTCGCCAAGACAAGACTATTCTATGCAAGTCAAGTTACACAGG 900

[illegible]

QY 1088 AAGATTATCTAGCTATGCACTTAATGATGTTGCCATGCTGTGACCACTGCCG 1147
 DB 961 AGGAATTTATCTGGCAATGCACTTGTATGATGCTATGCTGCGCAACCACTGCCAC 1020
 QY 1148 CCGTCCCTGCTCCGAAATATACCTCCCTCTTCCAGAAAGTTCCGTCGGCAGTGGA 1207
 DB 1021 CTGTCTGCTCCGAAATATACCTCCCTCTTCCAGAAAGTTCCGTCGGCAGTGGA 1080
 QY 1208 TGTCCGCTAATAGCTCTCTTCTGTGATCAGAGCTGCTGTGAGAGCCGTCGTCAAGG 1267
 DB 1081 TATCTGCTAATAGCTCAACATCTGTATGATCAGAGCTTACCAAGAGAAACGATTGGAAG 1140
 QY 1268 ATGAGCAGCAGC---CAGAGAGAACTGCTGTGATGATTTGAAGATTAAGACGGGAGA 1324
 DB 1141 ATGAGCAACAACAATTTGAAAGAAATTTACCTGTAAAGTTGAAAGTAAACCGGAGAG 1200
 QY 1325 ACTTCGAGCAGCAGCAGTGTGAGCTGTGAGAAAGCCGCCCAAGCGCTTTGGACAGCAGC 1384
 DB 1201 ACTTTGAACGTGGCAACCTGGAACTGGAGAAACCAAGGCAAGCTCTCTTGGAAACAGCAGC 1260
 QY 1385 GCAAAAGCAGAGCAGCTGTGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGG 1444
 DB 1261 GCAAGGAGCAGAGCGCTGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGG 1320
 QY 1445 AGCGCCAGAGCAGAGCGCAAGCGGAGCTGTGAGCTGTGAGAGAGCTGTGAGAGAGCAGC 1504
 DB 1321 AGCGCCAGAGCAGAGCGCAAGCGGAGCTGTGAGAGAGCTGTGAGAGAGCAGC 1380
 QY 1505 GGGAGCTGAGCAGCAGCAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGGAGCGG 1564
 DB 1381 GGGAGCTGAGCAGCAGCAGAGAGAGAGAGAGAGAGAGATTTGAGAGCGCGGAGCGG 1440
 QY 1565 CAAAACGGGAACTGTGAAAGCAGCAGCACTTGAATGGGAAACCGAACCGAGACAGAAAC 1624
 DB 1441 CAAAACGGGAACTGTGAAAGCAGCAGCACTTGAATGGGAAACCGAAACCGAAAC 1500
 QY 1625 TCCTGAATCAGAGAAACAAGAGCAGAGAGCGGAGCTGTGAGAGCGCAAGAGAGAG 1684
 DB 1501 TACTTAATCAAAAGAAACAAGAGCAGAGAGCACTAGTTGTAAGTGAAGAGAGAGAGAG 1560
 QY 1685 CTCTGAGTTGAGTTGAGAGCTGTGATGACAAAGAGCATGAGTGAAGAGAGAGCTTC 1744
 DB 1561 CTCTGAGTTGAGTTGAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATG 1620
 QY 1745 AGATATCAGAGTGTGAGCTGTGAGCAACCGAGAGAGAGATTTGAGAGCAGAGAGAGCTTA 1804
 DB 1621 AAGATATCAGAGTGTGAGCTGTGAGCAACCGAGAGAGAGATTTGAGAGCAGAGAGAGCTTA 1680
 QY 1805 GAGAGCTAAGAAATTTGAGAAATCAACCACTTACAGAGCAGAGTGTGAGAGAGATCTAGCAAA 1864
 DB 1681 GAGAGCTAAGAAATTTGAGAAATCAACCACTTACAGAGCAGAGATCTAGCAAA 1740
 QY 1865 TGTGTTGAAGACTTATTTCCAGAGAAACAGATCTCATGATGATGATGATGATGATGATGATG 1924
 DB 1741 TGTGTTGAAGACTTATTTCCAGAGAAACAGATCTCATGATGATGATGATGATGATGATGATG 1800
 QY 1925 AGAAGAGTTTGAATGAGAGCTGCTTCTTAACCTCAAAAGAGCTGTGAGAGAGAGAGAGC 1984
 DB 1801 AGAAGAGTTTGAATGAGAGCTGCTTCTTAACCTCAAAAGAGCTGTGAGAGAGAGAGAGC 1860
 QY 1985 TGGCCCGGAGCAGCTCCGAGAGCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 2044
 DB 1861 TGGCCCGGAGCAGCTCCGAGAGCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1920
 QY 2045 TGGAGAGAGTTGATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 2104
 DB 1921 TGGAGAGAGTTGATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1980
 QY 2105 AGCAACTTCAAGAGAGAGAGTCCCTGAGAGCAGCGCAGCTGAGAGAGAGAGAGAGAGAGAGC 2164
 DB 1981 AACCACTTCAAGAGAGAGAGTCCCTGAGAGCAGCGCAGCTGAGAGAGAGAGAGAGAGAGAGC 2040

QY 2165 GGAAGAGCCTGGAGTTAGAGAGCAAAAGAGAGA 2198
 DB 2041 GAAAGATCATGAATTTAGAGAGCAAAAGAGAGAGA 2074
 RESULT 3
 B0942708
 LOCUS
 DEFINITION
 B0942708 911 bp mRNA linear EST 21-AUG-2002
 AGENCOURT 8764676 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6314690
 5', mRNA sequence.
 B0942708
 B0942708.1 GI:22358186
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 911)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 CDNA Library Preparation: Resgen, Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM3743 row: n column: 03
 High quality sequence start: 10
 High quality sequence stop: 564.
 Location/Qualifiers
 1. 911
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6314690"
 /clone_1b="NIH_MGC_129"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: olfactory epithelium; Vector:
 pCMV-SPORT6.1.cdb; Site: 1; EcoRV; Site 2: NotI; Cloned
 unidirectionally. Primer: Oligo dt. Average insert size
 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this
 is a NIH MGC library."
 BASE COUNT 271 a 210 c 280 g 150 t
 ORIGIN
 Query Match 15.6%; Score 795.6; DB 14; Length 911;
 Best Local Similarity 98.0%; Pred. No. 4.3e-179;
 Matches 890; Conservative 0; Mismatches 9; Indels 9; Gaps 8;
 QY 936 GCAATTATTTCAACGCGACCAAACTATGAGTGAACATTTAAGCTCCAGGCAAG 995
 DB 1 GCAATTATTTCAACGCGACCAAACTATGAGTGAACATTTAAGCTCCAGGCAAG 60
 QY 996 AACTATTCTATGCAATCAAGTTTACCCGAGCTCAGCTGCTCAATATGATGATGATGATGATGATG 1055
 DB 61 AACTATTCTATGCAATCAAGTTTACCCGAGCTCAGCTGCTCAATATGATGATGATGATGATGATG 120
 QY 1056 TGACATTGATCAAGATGAAAACTCACTGCGAGAAATTTATCTAGCTATGCACTTAAT 1115
 DB 121 TGACATTGATCAAGATGAAAACTCACTGCGAGAAATTTATCTAGCTATGCACTTAAT 1175
 QY 1116 TGATGTTGCGATGCTGTGTCAGGCACTGCGCGCTGCTGCTGCGCAATATCTCTTC 1175
 DB 180 TGATGTTGCGATGCTGTGTCAGGCACTGCGCGCTGCTGCTGCGCAATATCTCTTC 239
 QY 1176 TTCTCTTCAAGAGAGTTGGCTCCGAGAGTGGAGTCCGTCATATGAGCTCTTCTCTGGA 1235
 DB 240 TTCTCTTCAAGAGAGTTGGCTCCGAGAGTGGAGTCCGTCATATGAGCTCTTCTCTGGA 299
 QY 1236 TCAGAGGCTGCTGAGAGCGCTGCTGATGAGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1295

```

|||||
Db 300 TCAGAGGCTGCTCAGAGAGCCGTCGTACAGGATGAGCAGACGCCAGAGAAAGAACTGCC 359
QY 1296 TGTGACATTTGAATTAAGAGCGGGAGAACTTCGAGCCAGGCGAGTGTGAGTGGAGAA 1355
Db 360 TGTGACATTTGAATTAAGAGCGGGAGAACTTCGAGCCAGGCGAGTGTGAGTGGAGAA 419
QY 1356 GCCCGCCCAAGCCCTCTTGGAGCAGCAGCGCAAAAGAGCAGAGCGGTTGGCTCAGCTGGA 1415
Db 420 GCCCGCCCAAGCCCTCTTGGAGCAGCAGCGCAAAAGAGCAGAGCGGTTGGCTCAGCTGGA 479
QY 1416 GCCCGCCAGCAGAGAGAGAGAGCGGAGCGCCGAGAGCAGAGAGCCAGAGCGGCTAGCT 1475
Db 480 GCCCGCCAGCAGAGAGAGAGAGCGGAGCGCCGAGAGCAGAGAGCGCAAGCGGCAAGCT 539
QY 1476 GAGCTGTGAGAACCTGAGAGAGAGAGAGCGGAGCGGAGCGGAGCGGAGAGAGAGAG 1535
Db 540 GAGCTGTGAGAACCTGAGAGAGAGAGAGCGGAGCGGAGCGGAGCGGAGAGAGAGAGAG 599
QY 1536 GAGAGAGAGATGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1595
Db 600 GAGAGAGAGATGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 659
QY 1596 TGAATGGAA-C-GGAACCGGAGACAGGAACTCTTGAATCAGAGAAACAGAGCAGAGAG 1654
Db 660 TGAATGGAA-C-GGAACCGGAGACAGGAACTCTTGAATCAGAGAAACAGAGCAGAGAG 719
QY 1655 GC--ACGCTGTCTTGAAGCAGAGAGAGAGAGCTGTGAGATTGAGTTGAAGCTCTGAA 1712
Db 720 GCCACCGGCTGCTTGAAGCAGAGAGAGAGAGCTGTGAGATTGAGTTGAAGCTCTGAA 779
QY 1713 TGACAAAAGCATCAGCTAGTAAGAGAAACTTCAGGATATCAGGT-GTCGATGGCAAGCC 1771
Db 780 TGACAAAAGCATCAGCTAGTAAGAGAAACTTCGAGATATCAGGTGCTCCATGGCAAGCC 829
QY 1772 AGA-GGCAAGAAATTGAGAGCAGC-AAAGCTTGA-GAGCTTAAGATTGC-TGAATC 1827
Db 840 AGAGGCGCAAGAAATTGAGAGCAGCAGAAACAAGTCTAAAGAGAGTAAAAATTGCTTGAATC 899
QY 1828 ACCCACTT 1835
Db 900 ACCCACTT 907

RESULT 4
LOCUS BM950476 732 bp mRNA linear EST 14-MAR-2002
DEFINITION UI-M-EHdp-buu-o-05-0-UI.r1 NIH_BMAP_EHdp Mus musculus cDNA clone
IMAGE:5687260 5', mRNA sequence.
ACCESSION BM950476
VERSION BM950476.1 GI:19434066
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers

```

```

source
1. 732
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5687260"
/clone_1db="NIH BMAP EHdp"
/tissue_type="Whole Brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/notes="Organ: brain; Vector: pYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GAGCAGAGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 208 a 172 c 201 g 148 t 3 others
ORIGIN
Query Match 14.2%; Score 722.8; DB 14; Length 732;
Best Local Similarity 99.3%; Pred. No. 1.1e-161;
Matches 724; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 3075 AGACATGTGTGTGTGTGAGAAATTCAAGTCAAGAGGTTGTTCCCAAGCTTACGT 3134
Db 4 ACATNNGTGTGTGTGTGAGAAATTCAAGTCAAGAGGTTGTTCCCAAGCTTACGT 63
QY 3135 GAACTCAATTCAGGCGCCGTAAGGAATTCACAAAGCATGATCTGGCCCTACTGAAAG 3194
Db 64 GAACTCAATTCAGGCGCCGTAAGGAATTCACAAAGCATGATCTGGCCCTACTGAAAG 123
QY 3195 TCCTGTAGTCTTAAGAGAGTGTCTCCCGGCGCCAGCCAGCATTCGCGAGAAAG 3254
Db 124 TCCTGTAGTCTTAAGAGAGTGTCTCCCGGCGCCAGCCAGCATTCGCGAGAAAG 183
QY 3255 GTTTATGTCATGTACACATACAGAGATTTGAGCAGAGAGATTTAACTTTACAGCAAG 3314
Db 184 GTTTATGTCATGTACACATACAGAGATTTGAGCAGAGAGATTTAACTTTACAGCAAG 243
QY 3315 GGATGTGATTGTGTACCAAGAAAGATGTGACTGTGAGCGGGAAGGTGGGCAACA 3374
Db 244 GGATGTGATTGTGTACCAAGAAAGATGTGACTGTGAGCGGGAAGGTGGGCAACA 303
QY 3375 GTCCGAGTCTTCCTTCTAATATGTAGAGGCTTAAAGTTCAGAGGCTCTGAACTGC 3434
Db 304 GTCCGAGTCTTCCTTCTAATATGTAGAGGCTTAAAGTTCAGAGGCTCTGAACTGC 363
QY 3435 TGGGAAAAAGGAGGATTTAGGAAAAAACTGAAATTGCGCAGATTATTCCTTACGC 3494
Db 364 TGGGAAAAAGGAGGATTTAGGAAAAAACTGAAATTGCGCAGATTATTCCTTACGC 423
QY 3495 TGTACTGTGTCCCGAACAATCACTCCGCTCTGGGAGCGATTCGATCGGAAAAA 3554
Db 424 TGTACTGTGTCCCGAACAATCACTCCGCTCTGGGAGCGATTCGATCGGAAAAA 483
QY 3555 GAAACCCAGTGTGTGAGAGAGAACTGCAAGCTCGAGGAAAAAGCGCCAGATAGG 3614
Db 484 GAAACCCAGTGTGTGAGAGAGAACTGCAAGCTCGAGGAAAAAGCGCCAGATAGG 543
QY 3615 GTGTGTTTCAGCAAAATTTATGTAACCTTTAAAGCCCGGAGAACAAATATACCCCAAC 3674
Db 544 GTGTGTTTCAGCAAAATTTATGTAACCTTTAAAGCCCGGAGAACAAATATACCCCAAC 603
QY 3675 TGAGCTACCCAGACCGGAGTGCAGCAGCATGTGCCAGGATCGGAGATACGATTA 3734

```


Db 604 TGAGCTACCCAGACCGCACTGACCGACGAGTGTGCAAGTATCGGATGTACGATTA 663

QY 3735 CACCGCCAGACGATGACGATCTAGCTTCAGCAAAAGGCGAGATCATCAAGCTCTCAA 3794

Db 664 CACCGCCAGACGATGACGATCTAGCTTCAGCAAAAGGCGAGATCATCAAGCTCTCAA 723

QY 3795 CAAGAGAGA 3803

Db 724 CAAGAGAGA 732

RESULT 5

LOCUS BM944544 738 bp mRNA linear EST 14-MAR-2002

DEFINITION UI-M-EHOP-bvr-j-08-0-UI.r1 NIH_BMAP_EHOP Mus musculus cDNA clone

ACCESSION IMAGE:5695975 5', mRNA sequence.

VERSION BM944544

KEYWORDS BM944544.1 GI:19428129

SOURCE EST.

ORGANISM house mouse.

Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 738)

TITLE NIH-MGC http://mgi.nci.nih.gov/

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pyx-5

FEATURES

Location/Qualifiers

1..738

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:5695975"

/clone_1ib="NIH_BMAP_EHOP"

/tissue_type="Whole Brain"

/dev_stage="embryo 18.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/note="Organ: brain; Vector: pyx-asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with Not I, and then cloned directionally into pyx-asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemm Chin, Ph.D., program coordinator."

BASE COUNT 198 a 213 c 184 g 143 t

ORIGIN

Query Match 14.2%; Score 721.2; DB 14; Length 738;

Best Local Similarity 99.5%; Pred. No. 2.5e-161;

Matches 734; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 2568 CCGAGCTGAGAGCCGATGCTTGGAGAGAGCTGAAGAGACCGGATGTTCCC 2627

Db 1 CCGAGCTGAGAGCCGATGCTTGGAGAGAGCTGAAGAGACCGGATGTTCCC 60

QY 2628 TGCAAACTATGACAGAAAAGATTCCAGAAAATGAGGTTCCCACTCCAGCCAAACAGTAGC 2687

Db 61 TGCAAACTATGACAGAAAAGATTCCAGAAAATGAGGTTCCCACTCCAGCCAAACAGTAGC 120

QY 2688 CGATCTGACATCTGCCCCCTGCCCCCAACTGCTCTGCTGAGACCCCTGCTCTCTTGGC 2747

Db 121 CGATCTGACATCTGCCCCCTGCCCCCAACTGCTCTGCTGAGACCCCTGCTCTCTTGGC 180

QY 2748 AGTACCTCTCTGAGCCCTCCCAACCCCAACCACTGGGCGAGCTTCAGTCCACGTC 2807

Db 181 AGTACCTCTCTGAGCCCTCCCAACCCCAACCACTGGGCGAGCTTCAGTCCACGTC 240

QY 2808 GCCCAGACGCTCAACAGAGAAAGCCAGAAAGGACCAACTGGGATATCGTGGCGGCTCAGCC 2867

Db 241 GCCCAGACGCTCAACAGAGAAAGCCAGAAAGGACCAACTGGGATATCGTGGCGGCTCAGCC 300

QY 2868 TTCTCTGACGCTACCTTGTGCTGGCCAGTTACGGCAGAGATCAGCCTTTACCCAGCCAC 2927

Db 301 TTCTCTGACGCTACCTTGTGCTGGCCAGTTACGGCAGAGATCAGCCTTTACCCAGCCAC 360

QY 2928 AGCCACTGGCTCTCCCATCTCCCGTCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2987

Db 361 AGCCACTGGCTCTCCCATCTCCCGTCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420

QY 2988 AGCCCAAGCCCTGTATCTCTGAGAGCCCAAAAAGACCAACACTTAATTTTAAACAAAG 3047

Db 421 AGCCCAAGCCCTGTATCTCTGAGAGCCCAAAAAGACCAACACTTAATTTTAAACAAAG 480

QY 3048 TGAGGTATACCGTCTGGAACAGCAAGACATGTGTGTTGGAGAACTTCAAGGTCA 3107

Db 481 TGAGGTATACCGTCTGGAACAGCAAGACATGTGTGTTGGAGAACTTCAAGGTCA 540

QY 3108 GAGGAGTGTGTTCCCAAGCTTACGTAACATTTTCAGGAGCCGCTTAAGAAATCCAC 3167

Db 541 GAGGAGTGTGTTCCCAAGCTTACGTAACATTTTCAGGAGCCGCTTAAGAAATCCAC 600

QY 3168 AAGCATGATTAATGAGCCCTTACTGAAAGTCTGCTAGTCTTAAGAGAGTGGCTTCCCGGC 3227

Db 601 AAGCATGATTAATGAGCCCTTACTGAAAGTCTGCTAGTCTTAAGAGAGTGGCTTCCCGGC 660

QY 3228 CGCCCAAGCCAGCCATTCGCGAGAGAGTTATTCATATACATACAGAGATCTCTGA 3287

Db 661 CGCCCAAGCCAGCCATTCGCGAGAGAG-TTATTCATATACATACAGAGATCTCTGA 719

QY 3288 GCAAGAGATTTAATCTT 3305

Db 720 GCAAGAGATTTAATCTT 737

RESULT 6

LOCUS BG922629 782 bp mRNA linear EST 05-JUN-2001

DEFINITION 60282158F1 NCI_CGAP_Mame Mus musculus cDNA clone IMAGE:4950524 5', mRNA sequence.

ACCESSION BG922629

VERSION BG922629.1 GI:14303105

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA library preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
plate: LLNL0905 row: e column: 21
High quality sequence stop: 716.

FEATURES

source

1..782

Location/Qualifiers

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4950524"
/clone_1b="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT

224 a 187 c 191 g 180 t

ORIGIN

Query Match 13.6%; Score 691.2; DB 13; Length 782;
Best Local Similarity 97.4%; Pred. No. 3.8e-154;
Matches 746; Conservative 0; Mismatches 13; Indels 7; Gaps 4;

3567 ATGCTGGAGAGAGAACTGCAAGCTGAGGAGAAAGCCCAATAGGCTGTTCCAGC 3626
1 ATGCTGGAGAGAGAACTGCAAGCTGAGGAGAAAGCCCAATAGGCTGTTCCAGC 60
3627 -AAATATGTCAACTTCTAAGCCCGGAGCAAGCAAAATCACCCTAGAGCTACCA 3685
61 GAAATATGTCAACTTCTAAGCCCGGAGCAAGCAAAATCACCCTAGAGCTACCA 120
3686 AGACCCGAGTGCAGCCAGCAGTGTGCGAGTGTGATTCAGATTACACCCGCCAGA 3745
121 AGACCCGAGTGCAGCCAGCAGTGTGCGAGTGTGATTCAGATTACACCCGCCAGA 180
3746 AGCAGTACGAACTAGCTTCAAGCAAGCCCAATCAGTCTTCAACAGAGAGACC 3805
181 AGCAGTACGAACTAGCTTCAAGCAAGCCCAATCAGTCTTCAACAGAGAGACC 240
3806 CGGACTGTGGAGAGAGAGAGTCAAGTGGGCAAGTGGGCTTCCCAATTAATGTA 3865
241 CGGACTGTGGAGAGAGAGTCAAGTGGGCAAGTGGGCTTCCCAATTAATGTA 300
3866 AGCTGACACAGACATGAGCCCGCAGCAGCAATCATATGTTGCCATCCCCCTC 3925
301 AGCTGACACAGACATGAGCCCGCAGCAGCAATCATATGTTGCCATCCCCCTC 360
3926 AGCTGACACAGATGAGCCCGCAGCAGCAATCATATGTTGCCATCCCCCTC 3985
361 AGCTGACACAGATGAGCCCGCAGCAGCAATCATATGTTGCCATCCCCCTC 420
3986 AGATGACAGCTTATCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 4445
421 AGATGACAGCTTATCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 480
4446 ACTGACAGAGAGAGTTTACCTGATGTCATGTCATGTCATGTCATGTCATGTCATG 4105
481 ACTGACAGAGAGAGTTTACCTGATGTCATGTCATGTCATGTCATGTCATGTCATG 540
4106 TCACTGCGTGCAGAGGAGAAATTCAGAACTGCAAGAGGAGTGGTCTCTTTGG 4165
541 TCACTGCGTGCAGAGGAGAAATTCAGAACTGCAAGAGGAGTGGTCTCTTTGG 600
4166 GCGTTTCTTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 4225
601 GCGTTTCTTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 657
4226 TAAGTATTTTAAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 4285

Db 658 TAAGTATTTTAAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 716
4286 TTGCTATTTTGGTTTAC-AAAAAGACCCTATCATGAGAGTGC 4329
717 TTGCTATTTTGGTTTAC-AAAAAGACCCTATCATGAGAGTGC 762

FEATURES

RESULT 7

B0941336

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 896)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-i@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2594 row: K column: 01
High quality sequence stop: 763.

FEATURES

source

1..896

Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6420600"
/clone_1b="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pGEMT; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT

214 a 232 c 230 g 216 t 4 others

ORIGIN

Query Match 13.0%; Score 660.2; DB 14; Length 896;
Best Local Similarity 86.8%; Pred. No. 9.9e-147;
Matches 787; Conservative 0; Mismatches 105; Indels 15; Gaps 5;

40 GCTTGTTGCTCCGTAAGTACGCGCGCTGCAAGAGATCCCAAGCGGCTCCGGAGC 99
1 GCTTGTTGCTCCGTAAGTACGCGCGCTGCAAGAGATCCCAAGCGGCTCCGGAGC 60
100 GCGGGAGGACAGAGGCGGCGGCGGAGTGTGTCGCGGCTGCGGACTCGGCTTC 159
61 -----GACAGAGAGGCGGCGGCGGAGTGTGTCGCGGCTGCGGACTCGGCTTC 109
160 CTCGC-GCGCGCTGCGGCGGCTGCAATTTGTGAGGGGCGGCGCGCAACCGCCCG 218
110 CTCGCAGGCGGCGGCGGAGGAGTGTGTCGCGGCGGCGCGCGGAGCGGCGCGG 169
219 GAGATGAGGCGGCTGATCAGCAAGGTGAACGTATGAACCATGCTGATTTCCACACC 278
170 GAGATGAGGCGGCTGATGCAAGGTGAACGTATGAACCATGCTGATTTCCACACC 229

QY 279 TTTCGGTGTAGGCTGATGCTGGGCAATACCTGTGAGGAAAGGCCAAGCATGACCA 338
 Db 230 TTTTGGTGGGAGCTGGATATCTGGGCAATACCTGTGAGGAAAGGCCAAGCATGACCA 289
 QY 339 GCAGTTCCTTAGGCTGAGAGCCGATAGCGGATTTATTCGCGGTACAGGAGCAACT 398
 Db 290 GCAGTTCCTTAGGCTGAGAGCCGATAGCGGATTTATTCGCGGTACAGGAGCAACT 349
 QY 399 TTTTTCCTAGCTGGGTATCTAGGCTGCTTCTTACGACAAATATGGCGCTAGCGACAT 458
 Db 350 TTTTTCCTAGCTGGGTATCTAGGCTGCTTCTTACGACAAATATGGCGCTAGCGACAT 409
 QY 459 GATTAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 518
 Db 410 GATTAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 469
 QY 519 GAAGCTAACAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 578
 Db 470 GAAGCTAACAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 529
 QY 579 GGGTATTTCCAGTGACACGACATTTGGTATAGAGGAGATTTCTAGCATGCCACCTAC 638
 Db 530 TGGTATTTCCAGTGACACGACATTTGGTATAGAGGAGATTTCTAGCATGCCACCTAC 589
 QY 639 AGCTGTGCTCCCTGCGCAATGGGCTCCATTCAGTGTGTGAATGTCTCCACCTTAGT 698
 Db 590 AGCTGTGCTCCCTGCGCAATGGGCTCCATTCAGTGTGTGAATGTCTCCACCTTAGT 649
 QY 699 ATCTTCTGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 758
 Db 650 ATCTTCTGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 709
 QY 759 GCGTCTGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 817
 Db 710 GCGTCTGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 769
 QY 818 GATCTGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 877
 Db 770 GATCTGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 829
 QY 878 CCAGCG-CCCCCTGCGCAATGGGCTCCCTGCGCAATGGGCTCCCTGCGCAATGGGCT 935
 Db 830 CCAGCGTCCCGCCAGTGGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 889
 QY 936 GCAGTTA 942
 Db 890 GCAGTTA 896

RESULT 8
 LOCUS BQ719508 979 bp mRNA linear EST 16-JUL-2002
 DEFINITION AGENCOURT 8219793 lupski_sympathetic_trunk Homo sapiens cDNA clone
 IMAGE:618811.1 5', mRNA sequence.
 ACCESSION BQ719508
 VERSION BQ719508.1 GI:21858405
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 979)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov
 Plate: LLM13583 row: c column: 24
 High quality sequence stop: 643.
 Location/Qualifiers
 1..979
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:618811.1"
 /issue_type="sympathetic trunk"
 /dev stage="adult", 16 yr
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORE (Life Technologies); Site 1:
 Note: Site 2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCAGCGCG-3' and
 5'-GACTAGTCTAGATGCGAGCGCGCGCT(15)-3'. Site selected >
 1 kb for average insert length 1.9 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine); available through Life
 Technologies."
 BASE COUNT 294 a 247 c 218 g 217 t 3 others
 ORIGIN
 Query Match 12.9%; Score 658.2; DB 14; Length 979;
 Best Local Similarity 84.2%; Pred. No. 3e-146;
 Matches 776; Conservative 0; Mismatches 141; Indels 5; Gaps 3;
 QY 549 ACTTCCCCCTGATGAGAAACAGACAGAGGCTATTTCCAGTGACACGACATTTGGTAT 608
 Db 1 ACTTCCCCCTGATGAGAAACAGACAGAGGCTATTTCCAGTGACACGACATTTGGTAT 60
 QY 609 AGAGGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668
 Db 61 GAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 QY 669 TCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 728
 Db 121 TCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 QY 729 CTTGCTAAGGAGGCTCCCTGCTATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 788
 Db 181 CTTGCTAAGGAGGCTCCCTGCTATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 QY 789 CACATGGCCAAAGATTTCTTCTTCAAGAGATGCTGCTCAAGGCTCAATTAACCTAA 848
 Db 241 CACATGGCCAAAGATTTCTTCTTCAAGAGATGCTGCTCAAGGCTCAATTAACCTAA 300
 QY 849 GTTACAGAGGCAATCATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908
 Db 301 GTTACAGAGGCAATCATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 QY 909 GCTTCACTCAATCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 968
 Db 361 GCTTCACTCAATCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 QY 969 TGGACACTTAACAGGCTCCCGAGCAAGAACTATTTCAAGCAATCAAGTTTACCAGGC 1028
 Db 421 TGGACACTTAACAGGCTCCCGAGCAAGAACTATTTCAAGCAATCAAGTTTACCAGGC 480
 QY 1029 TCAGCTGGCTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1088
 Db 481 TCAGCTGGCTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 QY 1089 AGAATTTATCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148
 Db 541 AGAATTTATCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 QY 1149 CGTCTGCTCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1208
 Db 601 TGTCTGCTCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660

TITLE
JOURNAL
COMMENT

M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takega, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
1. 650
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="D130043B15"
/clone_1ib="RIKEN full-length enriched, 12 days embryo spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site 1: Salt, Site 2: BamHI, cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'
GAGAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTT 3', cDNA was prepared by using trehalose chemo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'
GAGAGAGAGATTCGAGTTAATTAATTAATCCGCCGCCGCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified plasmid vector KS(+) after bulk excision from Lambda FLIC I."

BASE COUNT

178 a 180 c 169 g 122 t 1 others

Query Match 12.5%; Score 633.6; DB 10; Length 650;
Best local Similarity 99.2%; Pred. No. 2,2e-140;
Matches 636; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2463 GGAGAGTGTAAAGTGTATATATCCAGCGCTGTACCCCTTTGAAATCAGAAATCGCA 2522
Db 1 GGAGAGTGTAAAGTGTATATATCCAGCGCTGTACCCCTTTGAAATCAGAAATCGCA 60
Qy 2523 TGAATATCCATCCAGCCAGAGATATATGTCATGCTGTGATGAAAGCCAGATCGAGGCC 2582
Db 61 TGAGATCCATCCAGCCAGAGATATATGTCATGCTGTGATGAAAGCCAGATCGAGGCC 120
Qy 2583 AGATGCTGTGAGAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2642
Db 121 AGATGCTGTGAGAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy 2643 AAAGATTCAGAAATATGAGTTTCCACTCCAGCCAAACCGTGAACGATCTGACATCTGC 2702
Db 181 AAAGATTCAGAAATATGAGTTTCCACTCCAGCCAAACCGTGAACGATCTGACATCTGC 240
Qy 2703 CCCTGCCCCCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2762
Db 241 CCCTGCCCCCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Qy 2763 GCGCTCCACAAACCCCAACCACTGAGGAGACTTCAGTCCAGCTGCGCCAGAGCTCAAA 2822
Db 301 GCGCTCCACAAACCCCAACCACTGAGGAGACTTCAGTCCAGCTGCGCCAGAGCTCAAA 360
Qy 2823 CGAGAGCCAGAAACGAGCAACTGAGATAGTGGGCGGCTCAGCTTCTGACCCGATCC 2882
Db 361 CGAGAGCCAGAAACGAGCAACTGAGATAGTGGGCGGCTCAGCTTCTGACCCGATCC 420
Qy 2883 TAGTCTGCGCAGTTAGGCGAGAGATCAGCTTACCCCGACAGCCAGCTGCTCTC 2942
Db 421 TAGTCTGCGCAGTTAGGCGAGAGATCAGCTTACCCCGACAGCCAGCTGCTCTC 480
Qy 2943 CCCATCTCCGCTCTGCGCCAGAGGTGAAGAGGTGGAAGGCTACAGCCAGCCCTCTGA 3002
Db 481 CCCATCTCTCCGCTCTGCGCCAGAGGTGAAGAGGTGGAAGGCTACAGCCAGCCCTCTGA 540
Qy 3003 TCCCTGAGAGACCAAAAGAGCAACCACTTAATTTAAACAAAGTACGCTCATCCGT 3062
Db 541 TCCCTGAGAGACCAAAAGAGCAACCACTTAATTTAAACAAAGTACGCTCATCCGT 600
Qy 3063 TCTGAAACAGACAGCATGTGTGTTTGGAGAAATTCAG 3103
Db 601 TCTGAAACAGACAGCATGTGTGTTTGGAGAAATTCAG 641

RESULT 11
BQ898718
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ898718 1062 bp mRNA linear EST 16-AUG-2002
AGBNCOURT 8118759 Lupsaki dorsal root ganglion Homo sapiens cDNA
clone IMAGE:6180303 5', mRNA sequence.
BQ898718
BQ898718.1 GI:22290732
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1062)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@ncl.nih.gov
Tissue Procurement: Dr. James R. Lupsaki
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM3562 row: n column: 16
High quality sequence start: 116
High quality sequence stop: 760.

FEATURES

Location/Qualifiers
 1. 1062
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6180303"
 /clone_1ib="Lupski_dorsal_root_ganglion"
 /sex="male"
 /tissue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"
 /lab_host="DH10B"
 /note="vector: PCMV-SPORT6 (Life Technologies); Site_1:
 NotI; Site_2: SalI; cDNA made by oligo-dT priming.
 directionally cloned using the following adaptors:
 5'-TCGACCCGACGCTCCG-3' and
 5'-TCGACGCTTACATCGCGCGCGCCCT(15)-3'. Size selected >
 1 kb for average insert length 1.7 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life
 Technologies."

BASE COUNT 338 a 250 c 261 g 209 t 4 others

ORIGIN

Query Match 12.4%; Score 631.4; DB 14; Length 1062;
 Best Local Similarity 90.9%; Pred. No. 7.8e-140;
 Matches 715; Conservative 0; Mismatches 68; Indels 4; Gaps 4;

3242 TTCGCGGAGAGGTTTATGTCATGACATGAGAGTTGTGAGCAAGGATTAA 3301
 113 TTGGCGGAGAGGTTTATGTCATGACATGAGAGTTGTGAGCAAGGATTAA 172
 3302 CCTTTGAGCAAGGAGTGTGATTGTGTTACCAAGAAAGTGTGAGTGTGAGCAAGGAG 3361
 173 CCTTTGAGCAAGGAGTGTGATTGTGTTACCAAGAAAGTGTGAGTGTGAGCAAGGAG 232
 3362 CGGTGGGCGCAAGTCCGAGTCTTCTCTACTATGAGAGTTAAAGATTTCAGAG 3421
 233 CAGTGGGCGCAAGTCCGAGTCTTCTCTACTATGAGAGTTAAAGATTTCAGAG 292
 3422 GCTCTGGAATGCTGGGAAAACAGGAGTTAGAAAAAACTGAAATGGCCAGGTTA 3481
 293 GCTCTGGAATGCTGGGAAAACAGGAGTTAGAAAAAACTGAAATGGCCAGGTTA 352
 3482 TTGCTTCTACGCTGTACTGTGTCCTCCGACACTCACTGCTGCTGGCAGCTGATTC 3541
 353 TTGCTTCTACGCTGTGTCCTCCGACACTCACTGCTGCTGGCAGCTGATTC 412
 3542 TGATCCGAAAAGAACCCAGGTGATGTTGGGAGAGAGTGCAGACTCGAGGAAAA 3601
 413 TGATCCGAAAAGAACCCAGGTGATGTTGGGAGAGAGTGCAGACTCGAGGAAAA 472
 3602 AGGCGCGATAGAGGTGTTTCCAGCAATATGCAAACTTAAAGCCCGGAAACAGCA 3661
 473 AGGCGCGATAGAGGTGTTTCCAGCAATATGCAAACTTAAAGCCCGGAAACAGCA 532
 3662 AAATCACCCCACTGAGCTACCCAGAGCCGAGTGCAGCAGAGTGTGCGCAGTATCG 3721
 533 AAATCACCTCCAGAGAGCCACTTAAGTCAACAGCATTAGCGGCGAGTGTGCGCAGTATCG 592
 3722 GGATGTACGATTACCGCCCAAGACATGACCAACTAGCTTTCAGCAAGGCGCATGCA 3781
 593 GGATGTACGATTACCGCCCAAGACATGACCAACTAGCTTTCAGCAAGGCGCATGCA 652
 3782 TCAACGTCCTCAACAGAGAGAGCCGAGTGTGGAAGAGAAAGTCAAGTGGCAAGTTG 3841
 653 TCAACGTCCTCAACAGAGAGAGCCCTGACTGTGTGAAGAGAAAGTCAAGTGGCAAGTTG 712
 3842 GGCTCTTCCATCCCAATATATGTAAGTGCACACAGCATGAGCCCAAGCCAGCATGAA 3901
 713 GGCTCTTCCATCCCAATATATGTAAGTGCACACAGCATGAGCCCAAGCCAGCATGAA 772
 3902 TCATATGTTTCCATCCCGCTTCAAGCTTCTCAAGAGCCCAATATCCCAT 3961

Db 773 TCATATGTTTCCATCCCGCTTCAAGCTTCAAGAGCCCAATATCCCAT 832

Qy 3962 ATCACTGCGCC-AGAGGAGATGATGG-AGATGACGCC-TTGATCATGTGA-CTTGACGACAT 4017

Db 833 ATCACTGCGCCAAAGGATGATGGAGATGACGCCCTTGATCATGTGACCTTCAGCAT 892

Qy 4018 GATCACC 4024

Db 893 GATCACC 899

RESULT 12

AV246405

LOCUS

DEFINITION

AV246405 RIKEN full-length enriched, 0 day neonate head Mus
 musculus cDNA clone 4832408K23 3' similar to AF064243 Homo sapiens
 interseccion short form mRNA, mRNA sequence.

ACCESSION

AV246405 GI:16387096

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 4, 1999 this sequence version replaced gi:6233864.

CONTACT

Yoshihide Hayashizaki

LABORATORY

RIKEN Genomic

SCIENCE CENTER (GSC)

Yokohama Institute

THE INSTITUTE OF PHYSICAL

AND CHEMICAL RESEARCH (RIKEN)

TEL: 81-45-503-9222

Fax: 81-45-503-9216

E-MAIL: genome-res@gs.riken.go.jp

URL: http://genome-gsc.riken.go.jp/

CARNINCI, P., SHIBATA, Y.,

HAYASHI, Y., MURAMATSU, M.,

WAGI, K., FUJII, K., INOUE, K.,

OGAWA, K., TANAKA, T., MATSUURA

S., KAWAI, J., OKAZAKI, Y.,

MURAMATSU, M., INOUE, Y., KIRA, A. and

HAYASHIZAKI, Y.

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y.,

Shibata, K., Itoh, M., Carninci, P., Sugahara

Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A.,

Saito, T., Kiyosawa, H., Yamana, I., Aizawa

K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome-gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES		Location/Qualifiers
source	1..674	/organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="4832408K23" /clone_1ib="RIKEN full-length enriched, 0 day neonate head" /sex="mixed" /tissue_type="head" /dev_stage="0 day neonate" /lab_host="DH10B" /note="Site_1: SalI; Site_2: BamHI, cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATTCACAGAGCTCTTTTGTGTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAATTCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I."
BASE COUNT	150 a 153 c 152 g 219 t	
ORIGIN		
Query Match	12.1%;	Score 616; DB 9; Length 674;
Best Local Similarity	98.1%;	Pred. No. 3,5e-136;
Matches 655; Conservative	0;	Mismatches 10; Indels 3; Gaps 3;
4390	CTTTGCCGTTTGCCAGTGTCCCAACCAATGTGTAGTTGGGCGTGTCCTCCGCGT	4449
5	CCTTGCCCGTTGTCCAGTGTCCCAACCAATGTGTAGTTGGGCGTGTTCCCTGCGGT	64
4450	AGAGCAGAGAGATGAGTGAGTACCTGTTTGAAAATGTGTATGTAGACTGACCTGACT	4509
65	AGAGCAGAGAGAGATGAGTGAGTACCTGTTTGAAAATGTGTATGTAGACTGACCTGACT	124
4510	ATGGAAGGGGTTATGTTCTGTGTGACCACTACAGCTGACTGCGGCAATGTACATCTG	4569
125	ATGGAAGGGGTTATGTTCTGTGTGACCACTACAGCTGACTGCGGCAATGTACATCTG	184
4570	TACCGAAGATGAGTCTTCTCCATGAGTAAACCAACCAACCGGTGACAGTCTCTCATC	4629
185	TACCGAAGATGAGTCTTCTCCATGAGTAAACCAACCAACCGGTGACAGTCTCTCATC	243
4630	TACTGATTCATTTTACTTTGACACAGTACCTTTGAGCCACTGAGAAAGCAACCATGTT	4689
244	TACTGATTCATTTTACTTTGACACAGTACCTTTGAGCCACTGAGAAAGCAACCATGTT	303
4690	TCCGTTTGATCAGATGAGTACCTGAGTGGCCCGGTTTGTGTTTATTTTCAATCTGG	4749
304	TCCGTTTGATCAGATGAGTACCTGAGTGGCCCGGTTTGTGTTTATTTTCAATCTGG	363
4750	CATGTCCTTCACCACTAAATCTAGTAAGAGCCCACTGCCAGCGGTTACGATCATCACT	4809
364	CATGTCCTTCACCACTAAATCTAGTAAGAGCCCACTGCCAGCGGTTACGATCATCACT	423
4810	ACCCACCGCTTAGTCTCTGTTACGTAAGAGTTTATTCAGTGGCTTTTATGGAATATCT	4869
424	ACCCACCGCTTAGTCTCTGTTACGTAAGAGTTTATTCAGTGGCTTTTATGGAATATCT	483
4870	TGAACAGTAATCTTCTTGACAAAGAAAGATGTATGAAAGTCTCCCTGCAATTAATTTCC	4929
484	TGAACAGTAATCTTCTTGACAAAGAAAGATGTATGAAAGTCTCCCTGCAATTAATTTT-C	542
4930	CAGTGTTTACATTTTTTTAACTAGACTGTGAGGAGTGTCTACAGATTATATGAAATGGCC	4989

Db	543	CAGTGTTCACATTTTTCCTAGTACTGTG-6GGGTTTGCTACAGATTATATGAATGCGGC	601
QY	4990	TCCTGTCCTCCGTGTGTGCTTAACTTGTCCTGTAGCTGAAGCCGTGTCTTACATATTA	5049
Db	602	TCCTGTCCTCCGTGTGTGCTTAACTTGTCCTGTAGCTGAAGCCGTGTGTCTTACATATTA	661
QY	5050	GTTGGAG 5057	
Db	662	GTTGGAG 669	
RESULT	13		
LOCUS	BQ443095	610 bp	RNA linear EST 29-MAY-2002
DEFINITION	UI-M-EV0-bx-f-j-12-0-UI-r1 NIH_BMAP_EV0	Mus musculus	CDNA clone
ACCESSION	BQ443095		
VERSION	BQ443095.1	GI:21246207	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: GSGABs-remail.nih.gov Tissue Procurement: Dr. James Lin, University of Iowa CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa CDNA Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP) Seq primer: pYX-5. Location/Qualifiers 1. 610 /organism="Mus musculus" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:5707499" /clone_1lb="NIH_BMAP EV0" /tissue_type="whole brain" /dev_stage="embryo 15.5 dpc" /lab_host="DH10B (rti phage resistant)" /note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is GGGCGGTGAA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene discovery in the Developing Mouse Nervous System, supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."		
BASE COUNT	175 a	147 c	165 g 123 t
ORIGIN			
Query Match	11.9%;	Score 606.8;	DB 14; Length 610;
Best Local Similarity	99.7%;	Pred. No. 5.5e-134;	
Matches 608; Conservative	0; Mismatches 2;	Indels 0;	Gaps 0;
QY	3157	AGGAATTCACACAGATCGATCTACTGCGCCCTTACTGTAAGTCTCGTACTCTTAAAGAGACTG	3216


```

Db 1 AGGAATTCACAAGCATCATCTGCGCTTACGAAAGTCTGCTAGTCTAAAGAGAGTG 60
Qy 3217 GCTTCCCCGGCCCGCCCAAGCAGCATTTCCCGAGAAAGATTATTCATGACATAC 3276
Db 61 GCTTCCCCGGCCCGCCCAAGCAGCATTTCCCGAGAAAGATTATTCATGACATAC 120
Qy 3277 GAGAGTTCTGAGCAAGAGATTAACTTTACAGCAAGGGATGTGTTGTTACCAAG 3336
Db 121 GAGAGTTCTGAGCAAGAGATTAACTTTACAGCAAGGGATGTGTTGTTACCAAG 180
Qy 3337 AAAGATGTGACTGTGAGACGGGACGTTGGCGACAGTCCGGAGCTTCCCTTCTAAC 3396
Db 181 AAAGATGTGACTGTGAGACGGGACGTTGGCGACAGTCCGGAGCTTCCCTTCTAAC 240
Qy 3397 TATGTAGAGTTAAAGATTACAGAGGGCTCTGGAACCTGCGGAAAAACAGGAGTTAGGA 3456
Db 241 TATGTAGAGTTAAAGATTACAGAGGGCTCTGGAACCTGCGGAAAAACAGGAGTTAGGA 300
Qy 3457 AAAAAAAGTGAATTCAGGAGTTATGCTTCTTACGCTGCTACTGCTCCGACCACTC 3516
Db 301 AAAAAAAGTGAATTCAGGAGTTATGCTTCTTACGCTGCTACTGCTCCGACCACTC 360
Qy 3517 ACCCTGCTCTCTGGGACGCTGATTCTGATCCGAAAAAGAACCCAGGTGATGTGGAA 3576
Db 361 ACCCTGCTCTCTGGGACGCTGATTCTGATCCGAAAAAGAACCCAGGTGATGTGGAA 420
Qy 3577 GGAGAACTGCAACTCGAGGAGAAAAAGCCAGATAGGGGTGTTCCAGAAATTATGTC 3636
Db 421 GGAGAACTGCAACTCGAGGAGAAAAAGCCAGATAGGGGTGTTCCAGAAATTATGTC 480
Qy 3637 AAACCTTAAAGCCCGGAAACAAACAAATCACCCTGAGCTACCCAAAGCCGAGTG 3696
Db 481 AAACCTTAAAGCCCGGAAACAAACAAATCACCCTGAGCTACCCAAAGCCGAGTG 540
Qy 3697 CAGCCAGCACTGTGCGCAGTGATCGGATGTACGATTACACGCCCAAGCATGACGA 3756
Db 541 CAGCCAGCACTGTGCGCAGTGATCGGATGTACGATTACACGCCCAAGCATGACGA 600
Qy 3757 CTAGCTTCA 3766
Db 601 CTAGCTTCA 610

RESULT 14
LOCUS A1594919 603 bp mRNA linear EST 15-MAR-2000
DEFINITION mJ79910.Y1 Soares mouse p3NMf19.5 Mus musculus cDNA clone
sequence.
IMAGE:482370.5 similar to TR:042287 042287 INTERSECTIN.; mRNA
A1594919
ACCESSION A1594919
VERSION A1594919.1 GI:4603967
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mammalia (bases 1 to 603)
Marta, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Maira W/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LIND; contact the
IMAGE Consortium (info@image.lind.gov) for further information.

```

This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 MGI:293114
 Seq primer: -40RP from Gibco
 High quality sequence stop: 464
 POLYA=No.

FEATURES

source Location/Qualifiers
 1..603
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_id="IMAGE:482370"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT713D (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT713 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Mhoro Ko (Wayne State University)."
 BASE COUNT 175 a 144 c 168 g 115 t 1 others
 ORIGIN

Query Match 11.8%; Score 602; DB 9; Length 603;
 Best Local Similarity 99.8%; Pred. No. 7.6e-133;
 Matches 602; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 3292 GGAGATTAACTTTCAGCAAGGGATGTGTTGTTACCAAGAAAGTGTGACTGG 3351
Db 1 GGAGATTAACTTTCAGCAAGGGATGTGTTGTTACCAAGAAAGTGTGACTGG 60
Qy 3352 TGGACGGGAACGGTGGGCGACAGTCCGAGCTTCCCTTAACTATGTAGGCTTAA 3411
Db 61 TGGACGGGAACGGTGGGCGACAGTCCGAGCTTCCCTTAACTATGTAGGCTTAA 120
Qy 3412 GATTAGAGGGCTCTGGAACCTGCTGGGAAAAACAGGAGTTAGAAAAAAACCTGAATT 3471
Db 121 GATTAGAGGGCTCTGGAACCTGCTGGGAAAAACAGGAGTTAGAAAAAAACCTGAATT 180
Qy 3472 GCCCAGTTATGCTTCTTACGCTGCTACTGCTCCGAACTACCTGCTCTGGG 3531
Db 181 GCCCAGTTATGCTTCTTACGCTGCTACTGCTCCGAACTACCTGCTCTGGG 240
Qy 3532 CAGCTGATTCTGATCCGAAAAAGAACCCAGGTGATGTGGAAAGGAACTGCAAGCT 3591
Db 241 CAGCTGATTCTGATCCGAAAAAGAACCCAGGTGATGTGGAAAGGAACTGCAAGCT 300
Qy 3592 CGAGGAAAAAGCCGCAATAGGTGTTTCCAGCAATTATGTCAAACTCTAAGCCCC 3651
Db 301 CGAGGAAAAAGCCGCAATAGGTGTTTCCAGCAATTATGTCAAACTCTAAGCCCC 360
Qy 3652 GGAACAAGCAAAATCAGCCCACTAGCTTACCAAGACCGAGTGACAGCAGTGTGC 3711
Db 361 GGAACAAGCAAAATCAGCCCACTAGCTTACCAAGACCGAGTGACAGCAGTGTGC 420
Qy 3712 CAGGTGATGGGATGTACGATTACCGGCCCAAGACGATGACGAACTAGCTTACGCAA 3771
Db 421 CAGGTGATGGGATGTACGATTACCGGCCCAAGACGATGACGAACTAGCTTACGCAA 480
Qy 3772 GGCACATATTAACGTCCTTCAACAAGAGAGACCGGCTGTGGAAAGGAGAGTCACT 3831
Db 481 GGCACATATTAACGTCCTTCAACAAGAGAGACCGGCTGTGGAAAGGAGAGTCACT 540
Qy 3832 GGGCAAGTTGGGCTCTTCCCATCAATTATGTAAAGTGCACACAGATGAGACCCAGC 3891
Db 541 GGGCAAGTTGGGCTCTTCCCATCAATTATGTAAAGTGCACACAGATGAGACCCAGC 600

```


QY 3892 CAG 3894
Db 601 CAG 603

RESULT 15

LOCUS AM910367 606 bp mRNA linear EST 25-MAY-2000
DEFINITION ur80908.Y1 NCI CGAP Mam6 Mus musculus cDNA clone IMAGE:315638 5' similar to TR:Q920R4 Q920R4 ESHL PROTEIN. ;, mRNA sequence.
ACCESSION AM910367
VERSION AM910367.1 GI:8075608
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 606)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNI)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ULNI at: image.llnl.gov/image/html/resources.shtml

MG1:1059394
Seq primer: -40RP from Gibco
High quality sequence stop: 442.
Location/Qualifiers
1. 606

FEATURES
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:315638"
/clone_1ib="NCI-CCAP_Mam6"
/sex="female, virgin"
/tissue type="infiltrating ductal carcinoma"
/dev stage="5 months"
/lab host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 150 a 179 c 128 g 148 t 1 others
ORIGIN

Query Match 11.7%; Score 593.2; DB 10; Length 606;
Best Local Similarity 99.3%; Pred. No. 9.7e-131;
Matches 595; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 548 CACTTCCCTGTCATGAACAGCAACGAGTGTATTTCCAGTGCACGACATTTGGTA 607
Db 8 CGTTTCCCTGTCATGAACAGCAACGAGTGTATTTCCAGTGCACGACATTTGGTA 67
QY 608 TAGAGAGGATTTGCTAGATGCACCACTCAAGCTGTCTCTGTGCCAATGGGCTCA 667
Db 68 TAGAGAGGATTTGCTAGATGCACCACTCAAGCTGTCTCTGTGCCAATGGGCTCA 127
QY 668 TTCCAGTTGTGAATGTCTCCACCTTAGTATCTTCTCCCTCCAGAGAGAGTGCCTC 727
Db 128 TTCCAGTTGTGAATGTCTCCACCTTAGTATCTTCTCCCTCCAGAGAGAGTGCCTC 187
QY 728 CCCTGGCTAAGGGGCTCTCCCTCATACAGCTGTGCTGTGGTGGCATCTGCAG 787
Db 188 CCCTGGCTAAGGGGCTCTCCCTCATACAGCTGTGCTGTGGTGGCATCTGCAG 247

QY 788 CCACATGGCCAAAGAGTTCCTTCAGCAGATCTGTCTCAGGGTCAATTAACACTA 847
Db 248 CCACATGGCCAAAGAGTTCCTTCAGCAGATCTGTCTCAGGGTCAATTAACACTA 307
QY 848 AGTTACAGAAAGGCATCATTCATGTGATGTCGCGCGCCCTCCAGCAGCAATGGGCTG 907
Db 308 AGTTACAGAAAGGCATCATTCATGTGATGTCGCGCGCCCTCCAGCAGCAATGGGCTG 367
QY 908 TGCTCAGTCATCAAGGCTGAATATACAGCAGTATTAACAGCAGCAACAACTATGA 967
Db 368 TGCTCAGTCATCAAGGCTGAATATACAGCAGTATTAACAGCAGCAACAACTATGA 427
QY 968 GTGACACTTAACAGTCCCGCAGCAAGAACTATTCTCATGCAATCAAGTTACCCAGG 1027
Db 428 GTGACACTTAACAGTCCCGCAGCAAGAACTATTCTCATGCAATCAAGTTACCCAGG 487
QY 1028 CTCAGCTGGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGGAACCTGACG 1087
Db 488 CTCAGCTGGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGGAACCTGACG 547
QY 1088 AAGAATTTATCTAGCTATGACCTAATTTGATGTGGCATGTGCTGACGCACTGCCG 1146
Db 548 AAGAATTTATCTAGCTATGACCTAATTTGATGTGGCATGTGCTGACGCACTGCCG 606

Search completed: March 14, 2003, 11:05:19
Job time : 6154.75 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:40:05 ; Search time 8885.04 Seconds

(without alignments)
11929.307 Million cell updates/sec

Title: US-09-674-237A-2

Perfect score: 3642

Sequence: 1 atggtcagttccacacaccc.....tggaccaccagccagcatga 3642

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1.*
1: gb_bat.*
2: gb_hgt.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_ph.*
7: gb_pl.*
8: gb_pl.*
9: gb_pl.*
10: gb_ro.*
11: gb_ro.*
12: gb_sy.*
13: gb_un.*
14: gb_un.*
15: gb_vl.*
16: em_ba.*
17: em_fun.*
18: em_hum.*
19: em_in.*
20: em_mu.*
21: em_or.*
22: em_ov.*
23: em_ov.*
24: em_ov.*
25: em_pl.*
26: em_pl.*
27: em_pl.*
28: em_pl.*
29: em_pl.*
30: em_pl.*
31: em_pl.*
32: em_pl.*
33: em_pl.*
34: em_pl.*
35: em_pl.*
36: em_pl.*
37: em_pl.*
38: em_pl.*
39: em_pl.*
40: em_pl.*
41: em_pl.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3640.4	100.0	3723	10	AF132478
2	3639.4	99.9	5145	10	AF132481
3	3229.8	88.7	4025	10	AF127798
4	2831	77.7	5381	9	AF114488
5	2830	77.7	6439	9	AF114487
6	2827.8	77.6	5287	9	AF064243
7	2826.8	77.6	5287	9	AF064244
8	2816.2	77.3	3812	10	AF132672
9	2027.6	55.7	4103	5	AF032118
10	1259.2	34.6	2131	9	AK027846
11	986.4	27.1	3241	9	HS061166
12	749	20.6	1133	10	AF169621
13	721	19.8	5828	9	AF248540
14	721	19.8	5828	9	AB033082
15	713.8	19.6	4557	9	AF182199
16	713.8	19.6	6091	9	AF182198
17	677	18.6	4977	10	AF132479
18	673.8	17.8	3544	10	AF132480
19	648	17.8	4447	6	AX428899
20	531.4	14.6	78190	2	AC106760
21	531.4	14.6	179264	2	AC023112
22	477.4	13.1	2017	6	AX366305
23	387.4	10.6	1329	9	AK021545
24	342.2	9.4	1966	9	AF180522
25	332.4	9.1	1873	6	AF175271
26	331.4	9.1	1391	9	AF001630
27	331.4	9.1	2004	9	AF038189
28	327.8	9.0	171603	9	AC012629
29	258	7.1	747	6	AR175268
30	258	7.1	4053	9	HS061167
31	251	6.9	230097	2	AC126053
32	241.4	6.6	181782	2	AC116970
33	216.8	6.0	180637	2	AC098552
34	206.2	5.7	156481	2	AC123507
35	202.4	5.6	110000	9	AP000193
36	202.4	5.6	114929	9	AP000050
37	202.4	5.6	125242	9	AP000311
38	202.4	5.6	151516	9	AP000117
39	202.4	5.6	340000	9	AP001718
40	199.6	5.4	178801	2	AC121075
41	197.2	5.4	235426	2	AC121074
42	196	5.4	272545	2	AC090533
43	195.6	5.4	139113	2	AC129075
44	191.8	5.3	139113	2	AC129075
45	191.8	5.3	139113	2	AC129075

ALIGNMENTS

RESULT 1
AF132478
LOCUS AF132478 3723 bp mRNA linear ROD 09-MAR-1999
DEFINITION Mus musculus Esei protein mRNA, complete cds.
ACCESSION AF132478
VERSION AF132478.1 GI:4378884
KEYWORDS
SOURCE
ORGANISM
MUS MUSCULUS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 3723)
Sengar, A.S., Wang, W., Bishay, J., Cohen, S. and Egan, S.E.
The EH and SH domain Esei proteins regulate endocytosis by linking
TITLE
to dynamin and Eps15

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

EMBO J. 18 (5), 1159-1171 (1999)
99164083
10064583
2 (bases 1 to 3723)
Sengar, A.S., Wang, W., Cohen, S., Bishay, J. and Egan, S.E
Direct Submission
Submitted (02-MAR-1999) Programs in Cancer & Blood

FEATURES	Location/Qualifiers
source	1. .3723 /organism="Mus musculus" /db_xref="taxon:10090"
CDS	1. .3642

```

/codon_start=1
/product="Esel protein"
/protein_id="AAB19746.1"
/db_xref="GI:437885"

translation="MAQPFPTFGQSLDVMAITVEBAKHDOQFLSLKPIAGFITGDQAR
RNEFPQSGLPQVLAQIWMALMDNNMGDMVDVEISIAKKLKLKIQGLQVSTLPLPVMV
KOQVVALISAPAFGIQGISALMDPLTAAVPMGSIPLVVMGSPVLVSPVAPVAPVAPV
GAPVLIPLPAFPAHPAATPKSSFSRSFGPSQSNITKLOKQSDVADSPADQCKVTA
QSNLYKTYROLFNSHDKTMSGHLTGPOARITIMOSLLPAQALSIWNLSDIPDQKATVA
EEFLIAMHLIDVMSGQPLPVLPPEVYPPSFRVRVSSGMSVYSSSSVDRLPEEBS
SEDOQPEKKLPVTFEDKKRKNFRFSGVYELKRRKQALNBOQKQERLQALQERAEQER
KERROEOEARQOLELEKQERLEKQERLEKQERREERKELEREAKARELERQOLEMER
NRROELNORNKQEOGVTLKARRKTLEFEALNDKQLEGLQALRICATLOREED
IESNKSRELRITAEITIHQOLQESQMGILIKPEKQILSQLOVOONSILHPSLLIT
LKRALKEKELARQOLEBQLEVEVERETSKOEIDVNNOLKELEIHSKQOLQORSLI
EALTKQKQERQSKLELEKQEDQKQERQKQERQKQERQKQERQKQERQKQERQKQER
EDSRKKEAERBAERPMQDKQSRLEPHQOEPAKATQAPMSTFKGPIITLSAQBSKV
VYVYALYPPESRSHDEITIQGDIVMDESQTGGEPMGIGELKGTQMTGFPANVAKILIR
ENEPVTPAKPVTDLTISAPKRLARETPAPLPVTSSESTPNKMADFSFWPSSSENE
KPELEINDMTMAQPSLTLPSSAQORSAFPAATGSSPSPLVQGEQVEGLQAOALM
YPMKAKDNHNLNFKKSDVITVLBOODMMWFEQOGMGFPKSVYKLSIGSVRSTESI
DTGTGESPASIKRVASPAAKRAIRGEEIRVATYESSQGLTQOQGDIVATYTKDGD
MWGTGVDSKSVPPSPVNYRLKDSGSGTAGTGSGLGKKPBLAQVATYAAGPQBLTTL
APGQJLILIRKNPGGWMGEGELQARGKRRQIQGFPANVYKLTSPSTKSIITPELPIAV
QPAVCQVIGMYDYANDDELAFSKQIINVLNKNDEPDMWKGVSQGVGLFPSPVYKLT
ITMDPQSQ"

```

BASE COUNT	1065 a	940 c	1026 g	692 t
ORIGIN				

Query Match	100.0%	Score 3640.4	DB 10	Length 3723
Best Local Similarity	100.0%	Pred. No. 0		
Matches 3641	Conservative 0	Mismatches 1	Indels 0	Gaps 0

Qy	1	ATGGCTCAGTTTCCACACCTTTGGTGGTATGACCTGAAATGTCGGGACCAATCTGGAG	60
Db	1	ATGGCTCAGTTTCCACACCTTTGGTGGTACCTGGAATGTCGGGACCAATCTGGAG	60
Qy	61	GAAGGCGCAAGCATGACCAAGCTTCTTACCTGAAGCCGATAGCCGGAATTAATTA	120
Db	61	GAAGGCGCAAGCATGACCAAGCTTCTTACCTGAAGCCGATAGCCGGAATTAATTA	120
Qy	121	GGTATCAAGGAGAACTTTTTTCCAATCTGGGTTACTCAGCTGTCTTAGCAAA	180
Db	121	GGTATCAAGGAGAACTTTTTTCCAATCTGGGTTACTCAGCTGTCTTAGCAAA	180
Qy	181	ATATGGGGCTAGGGGACATGAATPACGATGGAAGGATGATCAAGTGGAAATTTTCCATA	240
Db	181	ATATGGGGCTAGGGGACATGAATPACGATGGAAGGATGATCAAGTGGAAATTTTCCATA	240
Qy	241	GCCATGAAGCTTATCAAACTGAAGTACAAAGATATACGCTCCCTCCACACTTCCCT	300
Db	241	GCCATGAAGCTTATCAAACTGAAGTACAAAGATATACGCTCCCTCCACACTTCCCT	300
Qy	301	GTCATGAACAGCAACCAAGTGGCTATTTCAGTGAACAGGATTTGGTATGAGGGATT	360
Db	301	GTCATGAACAGCAACCAAGTGGCTATTTCAGTGAACAGGATTTGGTATGAGGGATT	360
Qy	361	GCTAAGTGCACCACTCAGCTGTGCTCTGTGCCAATGGGCTCAATTCAGTTGTT	420
Db	361	GCTAAGTGCACCACTCAGCTGTGCTCTGTGCCAATGGGCTCAATTCAGTTGTT	420

Db	361	GCTACGATGACCAACCACTCACAGCTGTTGCTCTGTGGCAATTGGGGCTTCATTCAGTGT	42
Qy	421	GGAAAGTCTCCACCCCTTAAGTATCTTCTGTCCCTCCAGACAGACTGCTCCCTGGGTAC	480
Db	421	GGAAAGTCTCCACCCCTTAAGTATCTTCTGTCCCTCCAGACAGACTGCTCCCTGGGTAC	480
Qy	481	GGGGCTCTCCCTGTCATACAGCCCTGCTCCGCTTGGCGACTCTCAGCCACATGGCA	540
Db	481	GGGGCTCTCCCTGTCATACAGCCCTGCTCCGCTTGGCGACTCTCAGCCACATGGCA	540
Qy	541	AAAGATCTTCTCTTCAGACAGATCTGTGTCCAGGGTCACAATTAACTTAATTACAGAG	600
Db	541	AAAGATCTTCTCTTCAGACAGATCTGTGTCCAGGGTCACAATTAACTTAATTACAGAG	600
Qy	601	GCACATCATTCGATGTGTCCAGCGCCCTCCAGCAGCAGAAATGGGCTGGCTCAGTCA	660
Db	601	GCACATCATTCGATGTGTCCAGCGCCCTCCAGCAGCAGAAATGGGCTGGCTCAGTCA	660
Qy	661	TCGAAGCTGAATAACAGCAGATTATTCACAGCCACGACAAACTATGATGGACACTTA	720
Db	661	TCGAAGCTGAATAACAGCAGATTATTCACAGCCACGACAAACTATGATGGACACTTA	720
Qy	721	ACAGGTCCCCAGGGAAGAATCTATTCATACCAATTCAGATTACCCAGGCTCAGCTGGT	780
Db	721	ACAGGTCCCCAGGGAAGAATCTATTCATCAATCAAGTTTATCCCAAGGCTCAGCTGGT	780
Qy	781	TCATATATGGAATCTTCTTGAATTGATTCAGATGAGAAACTACTGCAAAATTTATC	840
Db	781	TCATATATGGAATCTTCTTGAATTGATTCAGATGAGAAACTACTGCAAAATTTATC	840
Qy	841	CTAGCTATGACACTTAATTGATGTTGCCATGTCTGTACACTGTGCGCTCTGCT	900
Db	841	CTAGCTATGACACTTAATTGATGTTGCCATGTCTGTACACTGTGCGCTCTGCT	900
Qy	901	CCAGAAATACATCCCTCTCTCTCCATCAGAAAGTTCGGTCCGGCAGTGGAGTGTCCGTCATA	960
Db	901	CCAGAAATACATCCCTCTCTCTCCATCAGAAAGTTCGGTCCGGCAGTGGAGTGTCCGTCATA	960
Qy	961	AGCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTCGTCAAGAGTATGAGCAGCAG	1020
Db	961	AGCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTCGTCAAGAGTATGAGCAGCAG	1020
Qy	1021	CCAGAGAAAGAACTGCTGTGACATTTGAAATATAAACAACGGGAGAACTTGCAGAGCAGC	1080
Db	1021	CCAGAGAAAGAACTGCTGTGACATTTGAAATATAAACAACGGGAGAACTTGCAGAGCAGC	1080
Qy	1081	AGTGTGAGCTGAGAAAGCCGCCCAAGGCTTTTGGACACACAGCGCAAAAGCAGCAG	1140
Db	1081	AGTGTGAGCTGAGAAAGCCGCCCAAGGCTTTTGGACACACAGCGCAAAAGCAGCAG	1140
Qy	1141	CGGTTGGCTCAGCTGAGAGCCGCCGACAGCAGAGAGAAAGACGGGAGTGCTCAGCAGCAG	1200
Db	1141	CGGTTGGCTCAGCTGAGAGCCGCCGACAGCAGAGAGAAAGACGGGAGTGCTCAGCAGCAG	1200
Qy	1201	GAGGCCAAGCGCAGCTGAGAGCTGTGAGAAAGCAGCTGGAACACAGCGGAGCTGTGAGCGG	1260
Db	1201	GAGGCCAAGCGCAGCTGAGAGCTGTGAGAAAGCAGCTGGAACACAGCGGAGCTGTGAGCGG	1260
Qy	1261	CAGCGAGAGAGAGAGAGAGAAAGAGATTCAGAGGCGCGAGGCGCGCAAAACGGGAACTG	1320
Db	1261	CAGCGAGAGAGAGAGAGAGAAAGAGATTCAGAGGCGCGAGGCGCGCAAAACGGGAACTG	1320
Qy	1321	GAAAGGAGCGACACTTGATATGGGAAACGAAACCGGAGCAGAAACTCTTGATATCAGG	1380
Db	1321	GAAAGGAGCGACACTTGATATGGGAAACGAAACCGGAGCAGAAACTCTTGATATCAGG	1380
Qy	1381	AACAAGAGCAGAGAGGGCACCTGTGTCTCTGAAGGCAAGAGAGAACTCTGGAGTTTAG	1440
Db	1381	AACAAGAGCAGAGAGGGCACCTGTGTCTCTGAAGGCAAGAGAGAACTCTGGAGTTTAG	1440
Qy	1441	TTGAGAGCTCTGAATGACAAAAAGCATCAGCTTGAAGGAAAACTTCAGAGATACAGGTGT	1500
Db	1441	TTGAGAGCTCTGAATGACAAAAAGCATCAGCTTGAAGGAAAACTTCAGAGATACAGGTGT	1500

QY 1501 CGACTGGCAACCCAGAGGCAAGAAATGAGAGCAGCAACAGTCTAGAGAGCTTAAGATT 1560
 DB 1501 CGACTGGCAACCCAGAGGCAAGAAATGAGAGCAGCAACAGTCTAGAGAGCTTAAGATT 1560
 QY 1561 GCTGAATTCACCACTTACAGCAGCAGTTCAGAGATCTTCAGCAATTCCTTGAAGACTT 1620
 DB 1561 GCTGAATTCACCACTTACAGCAGCAGTTCAGAGATCTTCAGCAATTCCTTGAAGACTT 1620
 QY 1621 ATTCCAGAGAAACAGATACCTAGACCACTTAACCAATCCAGAGAAACAGTTCGAT 1680
 DB 1621 ATTCCAGAGAAACAGATACCTAGACCACTTAACCAATCCAGAGAAACAGTTCGAT 1680
 QY 1681 AGAGACTCCGCTTCTTACCTCAAAAAGAGCTTGAAGCAAAAGAGTGGCCCGCAGCAG 1740
 DB 1681 AGAGACTCCGCTTCTTACCTCAAAAAGAGCTTGAAGCAAAAGAGTGGCCCGCAGCAG 1740
 QY 1741 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACCAAGTCAAAAGCTGCAAGAGATTGAT 1800
 DB 1741 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACCAAGTCAAAAGCTGCAAGAGATTGAT 1800
 QY 1801 GTTTTCAACCAACAGCTGAAGAGAACTGAGAGATACATAGCAAAACAGCAATTCAGAG 1860
 DB 1801 GTTTTCAACCAACAGCTGAAGAGAACTGAGAGATACATAGCAAAACAGCAATTCAGAG 1860
 QY 1861 CAGAGCTCCCTGAGGAGCGCGACTGAAGCAGAAAGAGAGAGAGAGAGCTTGAAG 1920
 DB 1861 CAGAGCTCCCTGAGGAGCGCGACTGAAGCAGAAAGAGAGAGAGAGAGAGCTTGAAG 1920
 QY 1921 TTAGAGAGAGCAAAAGAGAGAGCTCAGAGAGAGTTCAGAGAAAGAGCAAGATGGCTG 1980
 DB 1921 TTAGAGAGAGCAAAAGAGAGAGCTCAGAGAGAGTTCAGAGAAAGAGCAAGATGGCTG 1980
 QY 1981 GAGCATGTGACAG 2040
 DB 1981 GAGCATGTGACAG 2040
 QY 2041 AAG 2100
 DB 2041 AAG 2100
 QY 2101 GACAG 2160
 DB 2101 GACAG 2160
 QY 2161 CCCCTGCTTACCAAG 2220
 DB 2161 CCCCTGCTTACCAAG 2220
 QY 2221 GTATTATTACCGAG 2280
 DB 2221 GTATTATTACCGAG 2280
 QY 2281 CCAGAGAGATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 2340
 DB 2281 CCAGAGAGATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 2340
 QY 2341 GAGCTGAAG 2400
 DB 2341 GAGCTGAAG 2400
 QY 2401 GAGGTTCCCACTCCAG 2460
 DB 2401 GAGGTTCCCACTCCAG 2460
 QY 2461 GCTCTGCTGAG 2520
 DB 2461 GCTCTGCTGAG 2520
 QY 2521 AACAACTGGGAG 2580
 DB 2521 AACAACTGGGAG 2580

QY 2581 GACAACTGGGATACGTGGGCGGCTCAGCCTTCTCTGACCCGTAAGTCTAGTCTGCGCACTTA 2640
 DB 2581 GACAACTGGGATACGTGGGCGGCTCAGCCTTCTCTGACCCGTAAGTCTAGTCTGCGCACTTA 2640
 QY 2641 CGGAGAGATCAGCCTTTAACCCAG 2700
 DB 2641 CGGAGAGATCAGCCTTTAACCCAG 2700
 QY 2701 GGCAGAGGTGAAG 2760
 DB 2701 GGCAGAGGTGAAG 2760
 QY 2761 AAAGCAACCACTTAATTTTAAACAAAGTACAGTCAACCTGTTCTGAGAACACCAAGAG 2820
 DB 2761 AAAGCAACCACTTAATTTTAAACAAAGTACAGTCAACCTGTTCTGAGAACACCAAGAG 2820
 QY 2821 ATGTGTGTGTGTGAGAGAGTTCAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
 DB 2821 ATGTGTGTGTGTGAGAGAGTTCAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
 QY 2881 CTCATTTGAGGCGCCGTAAAGAGATCCAAAGATCGATCTGAGCCTTACGTAAGAGTCTT 2940
 DB 2881 CTCATTTGAGGCGCCGTAAAGAGATCCAAAGATCGATCTGAGCCTTACGTAAGAGTCTT 2940
 QY 2941 GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGCCATTTCCCGAGAGAGATT 3000
 DB 2941 GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGCCATTTCCCGAGAGAGATT 3000
 QY 3001 ATTGCAATGATACATACAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
 DB 3001 ATTGCAATGATACATACAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
 QY 3061 GTGATTTGTGTTTACCAAG 3120
 DB 3061 GTGATTTGTGTTTACCAAG 3120
 QY 3121 GGAATCTTCCCTTCTTACATGATGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
 DB 3121 GGAATCTTCCCTTCTTACATGATGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
 QY 3181 AAAACAGGAGAGTTTGAAG 3240
 DB 3181 AAAACAGGAGAGTTTGAAG 3240
 QY 3241 ACTGATCCCGAACAACCTCAGCCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
 DB 3241 ACTGATCCCGAACAACCTCAGCCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
 QY 3301 CCAAGTGTATGTTGGAG 3360
 DB 3301 CCAAGTGTATGTTGGAG 3360
 QY 3361 TTTCCAGCAAAATTAATGATCAAACTTCTTAAGCCCGGAGAACAGCAAAATCAACCCCACTGAG 3420
 DB 3361 TTTCCAGCAAAATTAATGATCAAACTTCTTAAGCCCGGAGAACAGCAAAATCAACCCCACTGAG 3420
 QY 3421 CTACCCAAAGCCGAGTGAAG 3480
 DB 3421 CTACCCAAAGCCGAGTGAAG 3480
 QY 3481 GCCCAGAGAGATGACGAATTAAGCTTCAAGAAAGCCAGATATCAAGTCTTCAACCAAG 3540
 DB 3481 GCCCAGAGAGATGACGAATTAAGCTTCAAGAAAGCCAGATATCAAGTCTTCAACCAAG 3540
 QY 3541 GAGGAGCCGGAAGTGTGAG 3600
 DB 3541 GAGGAGCCGGAAGTGTGAG 3600
 QY 3601 TATGTAAGCTGACCAAGAGATGAG 3660
 DB 3601 TATGTAAGCTGACCAAGAGATGAG 3660

RESULT 2
 AF132481 5145 bp mRNA linear ROD 09-MAR-1999
 LOCUS Mus musculus Esell protein mRNA, complete cds.
 DEFINITION AF132481
 ACCESSION AF132481.1 GI:4378890
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Mus musculus.
 Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 5145)
 Senger, A.S., Wang, W., Bishay, J., Cohen, S. and Egan, S.E.
 The EH and SH3 domain Ese proteins regulate cyclocytosis by linking
 to dynamin and Eps15
 EMBO J. 18 (5), 1159-1171 (1999)
 99164083
 10064583
 2 (bases 1 to 5145)
 Senger, A.S., Wang, W., Cohen, S., Bishay, J. and Egan, S.E.
 Direct Submission
 Submitted (02-MAR-1999) Programs in Cancer & Blood
 Research/Developmental Biology, The Hospital for Sick Children, 555
 University Avenue, Toronto, ON M5G 1X8, Canada
 Location/Qualifiers
 1. 5145
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 1. 5145
 /codon_start=1
 /product="Esell protein"
 /protein_id="AA19749.1"
 /db_xref="GI:4378891"
 /translation="MAQFPPTPGSGSIDVWAIIVVERAKHDQFLSLKPIAGTTGDQA
 RNFPGSGLPQVLAQIMALADNNDGMDQVEFSIAKLILKIQGQVLPSTPPVM
 KOQPAVISAAPAFIGIGIASMPPLTAVAPMGSIPIVGMSPISVPPAAPPALAN
 GAPVIOPLPAAPATWPKSSFSRSGSGSLNTKLOKQSPVAPAAEWAAP
 OSSRLKYRQLFNSHDKTSGHILGPARTILMOSLPOALQSVISSVSDORDPEPS
 SEDBOPEKKLVTFEDKRENFERSGSELEKRAQLLEQKKEBERLAQERAEQER
 KEREBOEAKROLEKOLEKORLELEREREREKLEERRAARLELERAEQERAEQER
 NEROELNORNEQEGTVLKARKTLEFELBALDKHOLEGKIDIRCLATQROE
 IESTNSRELRLAETIHLQOLOSQOMGRILPEKQILSDLOKVOONSILPDSILT
 LKRALEAKELARQOLEDOEVERETRSKLOEDIVNNOLKREIRHSKQLOKORSIL
 EEARLKQKEERKSLLEKQKDAORVRERDKOMLEHVOOEORPEKPEERIKR
 EDSVKKEAEERAKPEMODKSRLEPHRQERKLTQAPMSTTEGPIITISAQBSKY
 VYRRALYPRFESHDEITIQEDIVWDSQVGEKMGKGTGMPFPAVAKLP
 ENKEVTPAKVVDLTSAAPKULRETPALRETPATVTSSEPTTPNNADSESPWSSNE
 KEETPNWDTWAQPSLTVPSAQLRQSRAPATVTSGPSVPLQOGEKVGLOAL
 YPMRAKDNHNFNSKSDVITVLEQDMFMFGVQOKMPEKSYVKLISGVKRSKTSI
 DTGTPESPASLKRVASPAKPAIPGEELFAMVTYSSSQGDLTFQGGVIVVTKKDG
 WMTGTGSGVPSNVLKNDSESGTAGTSGIKGKPEIAVOIVASAAGSPDLTI
 ARGOLILRKNGPMWVEGELQARCKQOIGFPAHYVKLSISPGSKITTPLEUKTAN
 QPAVCVOTMTYITAKONDELAFSKQIILNVLNKDEPMWKEVSGQGLFPPSVKUL
 TDLMPDSQWMSDHLDMLEPTEKRRQGYHELIVTEENVVNDQLVTELFOKPLTE
 SLLLEKEVMAIFVWKEKLIMCNILKALRVRKMSSEKVPKMIIDILSAQPLHMQ
 PYIRFSCQILNGALAIQOKTEADPKEFEVKRLAMPCKGKPLSSFLIKSOAVTRY
 PLIKNILENTPENHPDHSHLKALEKAEELCSQVNEVREKENDRELEWIOAVOCE
 GLSBOLEVNSVNCIGPRKELHSGKLKYAKASKGLYGLFENDELILLOITPDLSSGT
 DKVPSKSNLOYKMYKTRPELNEUVKLPDPSGSEPIFHSIDRVYTLAASINER
 TAVOKIKASAEIYIETEKKREKAYLVRSQKATIGLIMNVVGEIILKCRSHGKS
 NPEYCVMTGSOCHITKTIQDTLNPKNWNSQCFIIDLQEVLCITVFERDQFSPDDFL
 GTEIRVTDIKKQDQSGKPVTKCLLHEVPTGEIIVVRLDLQLEFEP"
 BASE COUNT 1448 a 1298 c 1415 g 984 t
 ORIGIN
 Query Match 99.9%; Score 3639.4; DB 10; Length 5145;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 GAAAGGCGCAAGCATGACACAGAGTTCCTTAGCCTGAAGCCGATAGGGGATTTTACT 120
 DB 61 GAAAGGCGCAAGCATGACACAGAGTTCCTTAGCCTGAAGCCGATAGGGGATTTTACT 120
 QY 121 GGTGATCAAGCGAGAACTTTTTTCCAACTGSGGTTACCTCAGCCTGTCTTAGACAA 180
 DB 121 GGTGATCAAGCGAGAACTTTTTTCCAACTGSGGTTACCTCAGCCTGTCTTAGACAA 180
 QY 181 ATATGGGCGTAGGGAGATGAATAAAGATGAGAGATGATCAAGGAAATTTTCATA 240
 DB 181 ATATGGGCGTAGGGAGATGAATAAAGATGAGAGATGATCAAGGAAATTTTCATA 240
 QY 241 GCGATGAAGCTTATCAAACTGAAGCTCAAGAGATATGAGTCCCTCCACACTTCCCT 300
 DB 241 GCGATGAAGCTTATCAAACTGAAGCTCAAGAGATATGAGTCCCTCCACACTTCCCT 300
 QY 301 GTTCATGAAGACGCAACCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGGATT 360
 DB 301 GTTCATGAAGACGCAACCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGGATT 360
 QY 361 GCTAGCATGACCACTACAGCTGTTGCTCTGCTGCAATGGGCTCCATTCCAGTTGTT 420
 DB 361 GCTAGCATGACCACTACAGCTGTTGCTCTGCTGCAATGGGCTCCATTCCAGTTGTT 420
 QY 421 GGAATGCTCCACCTTATGATATCTTCTGTCCTCCAGCAGAGTGCTCCCTGGCTAAC 480
 DB 421 GGAATGCTCCACCTTATGATATCTTCTGTCCTCCAGCAGAGTGCTCCCTGGCTAAC 480
 QY 481 GGGGCTCTCCCGTCATACAGCCTCTGCTGCTGGTGGATCCTGAGCACAATGGCCA 540
 DB 481 GGGGCTCTCCCGTCATACAGCCTCTGCTGCTGGTGGATCCTGAGCACAATGGCCA 540
 QY 541 AAGAGTCTCTTCTTACAGAGATCTGATCCAGGTCACAAATTAACACTTAATACAGAG 600
 DB 541 AAGAGTCTCTTCTTACAGAGATCTGATCCAGGTCACAAATTAACACTTAATACAGAG 600
 QY 601 GCACAATCATTCGATGTCGCGCAGGCCCCCTCCAGCAGACAAATGGGCTGCTCAGTCA 660
 DB 601 GCACAATCATTCGATGTCGCGCAGGCCCCCTCCAGCAGACAAATGGGCTGCTCAGTCA 660
 QY 661 TCAAGGCTGAATTCAGAGCAAGTATTAACAGACGACACAAACTATGATGACACTTA 720
 DB 661 TCAAGGCTGAATTCAGAGCAAGTATTAACAGACGACACAAACTATGATGACACTTA 720
 QY 721 ACAAGTCCCCAGGCAAGACTATTTCTATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780
 DB 721 ACAAGTCCCCAGGCAAGACTATTTCTATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780
 QY 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATACTCAGCAAGAAATTTATC 840
 DB 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATACTCAGCAAGAAATTTATC 840
 QY 841 CTAGCTATGACCTAATATGATGTTGCCATGTGTCAGCAGCACTGCCCGCTCTGCTCT 900
 DB 841 CTAGCTATGACCTAATATGATGTTGCCATGTGTCAGCAGCACTGCCCGCTCTGCTCT 900
 QY 901 CCAATAAATATCCCTGCTCTTCAAGAGAGTTCGCTCCGCGAGTGGATATGCCGCTATA 960
 DB 901 CCAATAAATATCCCTGCTCTTCAAGAGAGTTCGCTCCGCGAGTGGATATGCCGCTATA 960
 QY 961 AGCTCTTCTCTGATGATCAAGAGGCTGCTGAGAGAGCCGCTCAAGAGATGAGCAGAG 1020
 DB 961 AGCTCTTCTCTGATGATCAAGAGGCTGCTGAGAGAGCCGCTCAAGAGATGAGCAGAG 1020
 QY 1021 CCAAGAGAGAACTGCTGTCATGATTTGAAGATTAAGAGGAGAACTTTCAGAGGAG 1080
 DB 1021 CCAAGAGAGAACTGCTGTCATGATTTGAAGATTAAGAGGAGAACTTTCAGAGGAG 1080
 QY 1081 AGTGTGAGCTGAGAGAGAGCCGCGCAAGGCTCTTGGAGCAGCAGCGCAAGAGCAGAG 1140
 DB 1081 AGTGTGAGCTGAGAGAGAGCCGCGCAAGGCTCTTGGAGCAGCAGCGCAAGAGCAGAG 1140

QY	1141	CGGTGGCTCAGCTGAGAGCCGGCCGAGCAAGAGAGGAAAGACCTGGAGCGCAGAGAG	1200
Db	1141	CGGTGGCTCAGCTGAGAGCCGGCCGAGCAAGAGAGGAAAGACCGGAGCGCAGAGAG	1200
QY	1201	GAGCCCAAGCGGCGAGCTGGAGCTTGGAAGCAGCTTGGAACACAGCGGAGCTGGAGCGG	1260
Db	1201	GAGCCCAAGCGGCGAGCTGGAGCTTGGAAGCAGCTTGGAACACACCGGAGCTGGAGCGG	1260
QY	1261	CAGCGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCGCGAATAACGGGAACTG	1320
Db	1261	CAGCGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCGCGAATAACGGGAACTG	1320
QY	1321	GAAGAGCAGAGGACAACTTGAATGGGAACGGAAACCGGAGACAGAACTCCTGAATCAGAGG	1380
Db	1321	GAAGAGCAGAGGACAACTTGAATGGGAACGGAAACCGGAGACAGAACTCCTGAATCAGAGG	1380
QY	1381	AACAAAGAGCAGAGAGGAGCAACCTGGTCTTGAAAGGCAAGAGAGAACACTCTGGAGTTGAG	1440
Db	1381	AACAAAGAGCAGAGAGGAGCAACCTGGTCTTGAAAGGCAAGAGAGAACACTCTGGAGTTGAG	1440
QY	1441	TTGAGAGCTCTGAATGACAAATAAGCATCAGCTTGAAAGGAAAACTTCAGATATCAGGTGT	1500
Db	1441	TTGAGAGCTCTGAATGACAAATAAGCATCAGCTTGAAAGGAAAACTTCAGATATCAGGTGT	1500
QY	1501	CGATGGCAACCCAGAGGCGAAAGAAATTGAGAGACGAACAACTGATAGAGCTGAAGAAATT	1560
Db	1501	CGATGGCAACCCAGAGGCGAAAGAAATTGAGAGACGAACAACTGATAGAGCTGAAGAAATT	1560
QY	1561	GCTGAATTCACCCACCTTACACGAGCAGATTGACAGAAATTCACAGCAATCTTGAAAGACTT	1620
Db	1561	GCTGAATTCACCCACCTTACACGAGCAGATTGACAGAAATTCACAGCAATCTTGAAAGACTT	1620
QY	1621	ATTTCAGAGAAACAGATACTCACTGACACAGTTAAAACAAAGTCAGCAGAAACAGTTTGAT	1680
Db	1621	ATTTCAGAGAAACAGATACTCACTGACACAGTTAAAACAAAGTCAGCAGAAACAGTTTGAT	1680
QY	1681	AGAGACTCTGCTTCTTACCTCTCAAAAAGAGCTTGGAAGCAAAAGAGCTGCCCGGACAG	1740
Db	1681	AGAGACTCTGCTTCTTACCTCTCAAAAAGAGCTTGGAAGCAAAAGAGCTGCCCGGACAG	1740
QY	1741	CTCCGGAGCAGCTGGGACGAGGTGGAGAGAGAGAACCAAGTCAAACTGACAGAAATTGAT	1800
Db	1741	CTCCGGAGCAGCTGGGACGAGGTGGAGAGAGAGAACCAAGTCAAACTGACAGAAATTGAT	1800
QY	1801	GTTTTCACAACCAAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGCAACTCCAGAG	1860
Db	1801	GTTTTCACAACCAAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGCAACTCCAGAG	1860
QY	1861	CAGAGGTCCTCTGAGGCGCAGCGCATCTGAAGCAAGAAAGACAGAGAGAGAGAGAGCTTGAG	1920
Db	1861	CAGAGGTCCTCTGAGGCGCAGCGCATCTGAAGCAAGAAAGACAGAGAGAGAGAGAGCTTGAG	1920
QY	1921	TTTAAAGAGCAAAAGAAAGACGCTCAAGAGACGAGTTCAAGAAAGGGAACAGCAATAGCTG	1980
Db	1921	TTTAAAGAGCAAAAGAAAGACGCTCAAGAGACGAGTTCAAGAAAGGGAACAGCAATAGCTG	1980
QY	1981	GAGCATGTGAGAGAGAGAGAGCAGCCGCCCGGAAAAACCCACAGAGAGAGAGAGACTG	2040
Db	1981	GAGCATGTGAGAGAGAGAGAGCAGCCGCCCGGAAAAACCCACAGAGAGAGAGAGACTG	2040
QY	2041	AAGAGGAGAGACAGTGTCAAGAAAGAGAGCGGAAGAGAGAGAGAGAGAGCCAGCCGGAATTGCA	2100
Db	2041	AAGAGGAGAGAGACAGTGTCAAGAAAGAGAGCGGAAGAGAGAGAGAGAGAGCCAGCCGGAATTGCA	2100
QY	2101	GACAGAGAGAGTGGCTTTTCCATCCGCATCAGAGACCAAGCTGAGCCACCCAGGCA	2160
Db	2101	GACAGAGAGAGTGGCTTTTCCATCCGCATCAGAGACCAAGCTGAGCCACCCAGGCA	2160
QY	2161	CCCTGCTCTACACAGAGAAAGCCCGCTTACATTTCTGCACAGAGAGGTGTAAATG	2220
Db	2161	CCCTGCTCTACACAGAGAAAGCCCGCTTACATTTCTGCACAGAGAGGTGTAAATG	2220
QY	2221	GTATATTACGAGCGCTGTACCCCTTTGAATCCAGAAAGTCAGATAGATCCATCCAG	2280

Db	2221	GTATATTACCGAGCGCTGTATCCCTTTTGAATCCAGAAAGTCAGATGAAATACCATCCAG	2280
Qy	2281	CCAGAGATATTAGTCATGTGTGATGAAAGCCAGACTGAGAGCCAGATGTCCTTGAGGA	2340
Db	2281	CCAGAGATATTAGTCATGTGTGATGAAAGCCAGACTGAGAGCCAGATGTCCTTGAGGA	2340
Qy	2341	GAGCTGAAAGGGAAGACGGGATGTCCTTCGCAAACTATGCAAGAAATAATTCAGAAAT	2400
Db	2341	GAGCTGAAAGGGAAGACGGGATGTCCTTCGCAAACTATGCAAGAAATAATTCAGAAAT	2400
Qy	2401	GAGGTTCCCACTCCAGCCAAACAGTAGCCGATCTGACATCGCCCTGCCCCCAACTG	2460
Db	2401	GAGGTTCCCACTCCAGCCAAACAGTAGCCGATCTGACATCGCCCTGCCCCCAACTG	2460
Qy	2461	GCTCTGCGTGAAGACCCCTGCTCCTTTGCCAGTGACCTCTTGAGCCCTCCACAACCCC	2520
Db	2461	GCTCTGCGTGAAGACCCCTGCTCCTTTGCCAGTGACCTCTTGAGCCCTCCACAACCCC	2520
Qy	2521	AACCACTGGGCAACTTCAGTTCCAGTGGCCCCAGCAGTCAAAAGAAACCCGAAACG	2580
Db	2521	AACCACTGGGCAACTTCAGTTCCAGTGGCCCCAGCAGTCAAAAGAAACCCGAAACG	2580
Qy	2581	GACAACTGGGATACGTGGGGGGCTCAGCCCTTCCTGACCCGTACTAGTGTGGCCAGTTA	2640
Db	2581	GACAACTGGGATACGTGGGGGGCTCAGCCCTTCCTGACCCGTACTAGTGTGGCCAGTTA	2640
Qy	2641	CGGAGAGATCAGCCTTTTACCCGAGCCAGCAGCAGTGGCTCTCTCCCATCTCCGTCCTG	2700
Db	2641	CGGAGAGATCAGCCTTTTACCCGAGCCAGCAGCAGTGGCTCTCTCCCATCTCCGTCCTG	2700
Qy	2701	GGCCAGGGTAAAAAGGTGGAAGGGCTTACAGCGCAAGCCCTGTATCTCTGGAGGCCAA	2760
Db	2701	GGCCAGGGTAAAAAGGTGGAAGGGCTTACAGCGCAAGCCCTGTATCTCTGGAGGCCAA	2760
Qy	2761	AAACACAACCACTTAAATTTTACAAAGTGACGTCATCCCGTCTGGAACAGCAAC	2820
Db	2761	AAACACAACCACTTAAATTTTACAAAGTGACGTCATCCCGTCTGGAACAGCAAC	2820
Qy	2821	ATGTGTGTGTTGGAGAGTTCAAGGTCAGAGGGTGTGTTCCCAAGTCTTACGTGAA	2880
Db	2821	ATGTGTGTGTTGGAGAGTTCAAGGTCAGAGGGTGTGTTCCCAAGTCTTACGTGAA	2880
Qy	2881	CTCATTTCAAGGCCCCGTAAAGAAATCCACAGATCATGATCTGSCCTTACTGAAGTCT	2940
Db	2881	CTCATTTCAAGGCCCCGTAAAGAAATCCACAGATCATGATCTGSCCTTACTGAAGTCT	2940
Qy	2941	GCTAGTCTTAAAGAGAGTGTTCCTCCGGCCGCAAGCCCATTCCTCCGGAGAAAGTTT	3000
Db	2941	GCTAGTCTTAAAGAGAGTGTTCCTCCGGCCGCAAGCCCATTCCTCCGGAGAAAGTTT	3000
Qy	3001	ATTGCCATGTACATACGAGAGTTCTGAGCAAGAGATTTTAACTTTACACAAAGGGAT	3060
Db	3001	ATTGCCATGTACATACGAGAGTTCTGAGCAAGAGATTTTAACTTTACACAAAGGGAT	3060
Qy	3061	GTAATTTGTGTTACCAAGAAATGTGTGACTGTGTGACGGGAACGGTGGCGCAAGTCC	3120
Db	3061	GTAATTTGTGTTACCAAGAAATGTGTGACTGTGTGACGGGAACGGTGGCGCAAGTCC	3120
Qy	3121	GGAATCTTCCCTTCTAATCTATGTGAGGCTTTAAAGATTCAAGAGGCTCTGGAATCTGTGG	3180
Db	3121	GGAATCTTCCCTTCTAATCTATGTGAGGCTTTAAAGATTCAAGAGGCTCTGGAATCTGTGG	3180
Qy	3181	AAAACAGGAGTTTAGAAAAAAACCTGAAATTGCCCAGGTTATTTGCTTCTCAACGCTCT	3240
Db	3181	AAAACAGGAGTTTAGAAAAAAACCTGAAATTGCCCAGGTTATTTGCTTCTCAACGCTCT	3240
Qy	3241	ACTGTGTCGGAACAATCAACCTGCTCTCTGGGAGCTGATTTCTGATCCGGAAAAAGAAC	3300
Db	3241	ACTGTGTCGGAACAATCAACCTGCTCTCTGGGAGCTGATTTCTGATCCGGAAAAAGAAC	3300
Qy	3301	CCAGTGGATGTGTGGGAAGAACTGCAAGTCCAGGGAAAAAACGCCAAGTAGGGTGG	3360

Qy	541	AAAGTCTTCCTTGACAGAGATCTGGTCCAGGGTCACAAATTAAACATGATTTACAGAG	600
Db	809	AAGGTTCTTCCTTTAGTAGATCTGGTCCAGGGTCACAACTPAAACATTAATTACAAAG	868
Qy	601	GCACAAATCATTTGCATGTCCGACGGCCCTCCAGCAGCAGATGGGCTGTCCCTCACTCA	660
Db	869	GCACAGTCATTTATGTGGCCAGTGTCCACCAAGTGGCAGAGTGGGCTGTTCCCTCACTCA	928
Qy	661	TCAAGGCTGAAATACAGGCAGTTATTCAACAGCCACGACAAAATTATGAGTGGACACTTA	720
Db	929	TCAAGACTGAAATACAGGCAAATATTCAATGACTGTGACAAAATTATGAGTGGACACTTA	988
Qy	721	ACAGGTCCTCCAGGCAGAACTATTCTCATGCAATCAAGTTTAACTCCAGGCTCACTGGCT	780
Db	989	ACAGGTCCTCCAGGCAGAACTATTCTTATGCAAGTGTATCCACAGGCTCACTGGCT	1048
Qy	781	TCAATATGGAATCTTTCTGTGACATTGATCAAGATGGAATACTCACTGCAGAAATTTATC	840
Db	1049	TCAATATGGAATCTTTCTGTGACATTGATCAAGATGGAATACTTACAGCAGAGAAATTTATC	1108
Qy	841	CTAGCATATGCACCTTAATTGATGTGGCATGTCTGGTCAGCCACTGCGCCGCTCTGCT	900
Db	1109	CTGGCAATGCACCTCAATTGATGTAGCTATGTCTGGCCAACTCCACTGCTCTGCT	1168
Qy	901	CCAGAAATACATCCCTCCTTCTTCAAGAAAGTTTCGCTCCGAGATGGGATTCCTGTATA	960
Db	1169	CCAGAAATACATCCACTTCTTTTAAAGAGTTTCATCTGGCAGTGGTATTCGTGATA	1228
Qy	961	AGGCTTCTTCTGTGGATTCAGAGGCTGCTGAGAGCCGCTGTCAGAGATATGACAGAG	1020
Db	1229	AGCTTCAATCTGTAGATCAGAGGCTTACAGAGAACCATGTTTATGAAATATGAACAA	1288
Qy	1021	C---CAGAGAAAGAACTGCCCTGTGACATTTGGAAGTAAAGACGGGAGAACTTCAAGCA	1077
Db	1289	CAATTAGAAAAGAAATTACTCTGTATACGTTTGAAGATTAAGAACGGGAGAACTTTGAAGCT	1348
Qy	1078	GGCAGTGTGAGCTGTGAGAAAGCCGCCCAAGGCTTTTGGACAGCAGCAGCAAGAGCAG	1137
Db	1349	GGCAACTCTGAACTGTGAGAAACCAAGGCAAGCTCTCTGTGGAACACACGCGAAAGAGCAG	1408
Qy	1138	GAGCGATTGGCTCAGCTGTGAGCGCGCCGACAGAGAGGAAAGACCGGAGAGCGCCAGAG	1197
Db	1409	GAGGCTTGGCCAGCTGTGAGCGGGCGAGACAGAGAGGAAAGACCTGTAGCGCCAGAG	1468
Qy	1198	CAGGAGGCCAAAGCGGCGAGCTGTGAGAGCAGCTTGAAACACAGCGGAGAGCTTGAG	1257
Db	1469	CAAGAGCGCAAAAAGACAACTGTGAACTGTGAGAAAGCACTGGAAGAACAGCGGGAGCTGAA	1528
Qy	1258	CGGCAAGCAGAGAGAGAGAGAGAGAAAGAGATGTAGAGGCGCGAGCGCCGAAACGGGAA	1317
Db	1529	CGGCAAGAGAGAGAGAGAGAGAGAAATTTGAAGGCGAGAGGCTGTCAAAACGGGAA	1588
Qy	1318	CTGAAAGGCAAGCGACAACTTGAATGGGAAACGGAACCGGAGACAGAACTCCTGATAG	1377
Db	1589	CTTAAAGGCAACGACAACTTGTAGTGGAAACGGAATCGAAGGCAAGAACTATTAATCAA	1648
Qy	1378	AGGAACAAGAGAGAGAGGCGCACCGTGTCTTGAAAGCAGAGAGGAAGACTCTGGAGTTT	1437
Db	1649	AGGAACAAGAGAGAGAGAGCATATGTTGTACTGAAGCAAAAGAAAAGACTTTGGAATTT	1708
Qy	1438	GAGTTAGAGCTCTGAATGACAAAGAACATCAGCTATGAAGAAAACCTTCAGATATTCAGG	1497
Db	1709	GAAATTAGAGCTCTTAATGTATTAAGACATCACTGAAAGGGAACCTTCAGATATTCAGA	1768
Qy	1498	TGTGCACTGGCAACCCAGAGGCAAGAAATTGADAGACGAAACAAGCTTATGAGAGCTAAGA	1557
Db	1769	TGTGCAATTGACACCCAAAGGCAAGAAATTTGAAGACAAACAATCTATGAGAGTTTGA	1828
Qy	1558	ATTGCTGAATCACCCACTTACAGCAGCAGATTTCAGAAATCTCAGCAATGCTTTGGAAGA	1617
Db	1829	ATTGCGCAATCACCCACTCTACAGCAACAATTACAGGAATCTCAGCAATGCTTTGGAAGA	1888
Qy	1618	CTTATTCAGAGAAACAGATATCTCAGTACCAAGTTAAAAACAGTCCAGCAACAAGTTTG	1677

Db	1889	CTTATTCGAGAAAACAGATATCTCATATGACCAATTTAAACAAGTTCAAGACAAACGTTTG	1948
QY	1678	CATAGACTCGCTTCTTACCTCAAAAGAGCTTTGAAACAAAGAGCTGGCCGGCAG	1737
Db	1949	CACAGATTTCACTTGTTCACACTTAAAGAGCTTTAGAAAGCAAAAGAACTGCTGGCAG	2008
QY	1738	CAGTCCGGAGAGCTGACACAGAGGTGGAGAGAGACAGAGTCAAACTCAGAGATT	1797
Db	2009	CACCTACGAAACCACTGATGAGAGGAGAAAGAACTAGTCAAACTACAGAGATT	2068
QY	1798	GATGTTTCAACACAGCTGAAGAACTGAGAGATACATAGCAAAACAGCACTCCAG	1857
Db	2069	GATATTTTCAATTAATCAGCTGAAAGAACTTAAGAGAAATACAAATAAACCAACTCCAG	2128
QY	1858	AAGCAGAGTCCCTGGAGCAGCGCATCTGAAGCAGAAAGACAGAGAGAGAGAGCTTG	1917
Db	2129	AAGCAAAAGTCCATGAGAGGCTGAAACACTGAAACAGAAAGAAACGAAAGTACTTA	2188
QY	1918	GAGTTAAGAAAGCAAAAGAAAGCGCTCAGACACAGATTTCAGSAAAGGACAGCAATGG	1977
Db	2189	GAATTAAGAAAAACAAAAAGAAAGGCCAAAGACGAGCTCAGAAAGAGGACAGAGATGG	2248
QY	1978	CTGAGAGATGTCAGCAGCAGAG--GAGCAGCACAGCCCCCGGAAACCCACGAGAGAGAC	2034
Db	2249	CTGAGCATGTGCAGCAGAGAGACAGCATCAGAGACCAAGAAACTCAGAGAGAGAA	2308
QY	2035	AGACTGAAGAGGAGACAGCTGTCAAGAAAGAGAGCGGAGAGAGAGCCAGCCGGA	2094
Db	2309	AAACTGAAAAAGGAGGAGAGTGTCAAAAGAAAGATGGCGAGAAAGCAACAGGAA	2368
QY	2095	ATGCAGACCAAGCAGAGTGGGCTTTTCATCCGCATCAGAGACCGCTTAAGCTGGCCAC	2154
Db	2369	GCAACAAACAGAGCTGGGTGGCTTTTCATCAACACAAAGAACAGCTPAAGCAGCTGTG	2428
QY	2155	CAGGACCCCTGGTGTACACAGAGAAAGGCGGCTTACCATTTCTGCACAGAGAGTGA	2214
Db	2429	CAGGACCCCTGGTCTACTGCAGAAAAAGTCCACTTACATTTCTGCACAGSAAATGTA	2488
QY	2215	AAAGTGATATTACCGAGGCGCTGACCCCTTTGAAATCCAGAAAGTCAGATGAGATCAC	2274
Db	2489	AAAGTGATATTACCGGAGCATGTACCCCTTTGAAATCCAGAAAGTCAGATGAGATCACT	2548
QY	2215	ATCCAGCCAGGATATATGTCAT-----GTTGATGAAAGCCAGTCTGGA	2319
Db	2549	ATCCAGCCAGGAGCATATGTCATGTTAAAGGGGAAATGGTGTGATGAAAGCCAAACTGGA	2608
QY	2320	GAGCAGGATGGCTTGGAGGAGCTGAAAGGGAACAGCGGATGGTCCCTCCTCAACTAT	2379
Db	2609	GAAACCGGCTGGCTTGGAGGAGAAATTAAGAGAAAGACAGGGTGGTCCCTGTCAAACTAT	2668
QY	2380	GCAGAAAAGATTCCAGAAAATGAGGTTCCCATCTCCAGCCMAACCAAGTACCGATCTGACA	2439
Db	2669	GCAGGAAAAATCCAGAAAATGAGGTTCCCGGCTCCAGTGAAACCAAGTACTGATTCACAA	2738
QY	2440	TCTGCCCTTCGCCCAATCTGGCTCTGCGTGAGACCCCTGCTCTTTGCGAGTACCTCT	2499
Db	2729	TCTGCCCTTCGCCCAATCTGGCTCTGCGTGAGACCCCGCTTTGGCAGATTAACCTCT	2788
QY	2500	TCTAGCGCTTCACAAACCCCAACACTGGGGCAGATTCAAGTTCCACGTTGGCCGACAGC	2555
Db	2789	TCAGAGCCCTTCACAGACCCCTTAATACCTGGGCGCACTTCAAGCTCAGCTGGGCCACAGC	2844
QY	2560	TCAAAACGAGAGCCAGAAACGGAACAATGGGATACGTGGGCGGCTCAGCTTCTGTGAC	2615
Db	2849	ACGAATAGAGAAACAGAAAGGATTAATCTGGGATGATGAGGACAGCCAGCCCTCTCTACAC	2908
QY	2620	GTACTAGTCTGCTGCAGTTTACGGCAGAGATCAGCTTTTATCCCAAGCCACAGCACTGGC	2673
Db	2909	GTTCCAAATGCGCGCAGTTTAAGGACAGAGGTCCGCTTTTATCTCAGCCACCGCACTGGC	2968
QY	2680	TCTCTCCCATCTTCGCTGGGCTGGGCGAGGGTGAAAGGTGAAAGGGGTACAAACCGAAGCC	2739

Dh 269 ATGCTCACTTTTCCAAACACCTTTTGGTGGCAGCTGGATATCTGGGCCATTAACCTAGAG 328
Qy 61 GAAAGGGCCAGCATGACCGACAGATTCTTAGCCTGAAGCCGATAGCGGATTTATCT 120
Dh 329 GAAAGGCGAAGCATGATCAGAGCTTCCATAGTTTAAAGCAATATCTGGATTCTACT 388
Qy 121 GGTGATCAAGCGAGAACTTTTTTTCATCTGGTTCCTCAGCTCTCTTAGACAA 180
Dh 389 GGTGATCAAGCGAGAACTTTTTTTCATCTGGTTCCTCAGCTCTCTTAGACAA 448
Qy 181 ATATGGGCGCTGCGGACATGATACATGAGAGATGATCAATGGATTTTTCATA 240
Dh 449 ATATGGGCGCTGCGGACATGATACATGAGAGATGATCAATGGATTTTTCATA 508
Qy 241 GCCATGAACCTTATCAAACTGAAGTACAAAGATATCAGCTCCCTCCACACTTCCCT 300
Dh 509 GCTATGAACCTTATCAAACTGAAGTACAAAGATATCAGCTCCCTCCACACTTCCCT 568
Qy 301 GTGATGAACGACGACCACTGCTATTTCCAGTGCACGACATTTGGTATGAGAGGAT 360
Dh 569 GTGATGAACGACGACCACTGCTATTTCCAGTGCACGACATTTGGTATGAGAGGAT 628
Qy 361 GCTAGATGCGACCACTGCACGCTGTTGCTCCGTCGCAATGGGCTCATTCAGTGT 420
Dh 629 GCGAGATGCGACCGCTGCACGCTGTTGCTCCAGTGCATGGGATCTCATCGATGTT 688
Qy 421 GGAATGCTCCACCTTAGTATCTGTCCTCCAGCAGCAGTGGCTCCCTGGCTAAC 480
Dh 689 GGAATGCTCCACCTTAGTATCTGTCCTCCAGCAGCAGTGGCTCCCTGGCTAAC 748
Qy 481 GGGGCTCTCCGCTCATPACAGCTCTGCTGCGTTGGCCTCTGACCCACATGGCCA 540
Dh 749 GGGGCTCTCCGCTCATPACAGCTCTGCTGCGTTGGCCTCTGACCCACATGGCCA 808
Qy 541 AAGAGTCTCTCTGAGCAGATGCTGGTCCAGGGTCACTTAAACCTAAGTTTACGAAG 600
Dh 809 AAGAGTCTCTCTGAGCAGATGCTGGTCCAGGGTCACTTAAACCTAAGTTTACGAAG 868
Qy 601 GCACATCAATTCGATGCTGCGCAGCGCCCTCCAGCAGCAGAAATGGGCTGGCTCAGTCA 660
Dh 869 GCACATCAATTCGATGCTGCGCAGCGCCCTCCAGCAGCAGAAATGGGCTGGCTCAGTCA 928
Qy 661 TCAGGCTGAATATCAGGCAAGTATTCAACAGCCAGCAGCAAACTATGAGTGAACATTTA 720
Dh 929 TCAGGCTGAATATCAGGCAAGTATTCAATAGTCATGACAAAATCTATGAGTGAACATTTA 988
Qy 721 ACGAGTCCCGCAGCAGAACTATTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780
Dh 989 ACGAGTCCCGCAGCAGAACTATTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 1048
Qy 781 TCATATGGAATTTCTGTCATTTGATCAAGATGAGAAAATCTCAGTCAAGAAATTTATC 840
Dh 1049 TCATATGGAATTTCTGTCATTTGATCAAGATGAGAAAATCTCAGTCAAGAAATTTATC 1108
Qy 841 CTAGCTATGCACTTAATGATGTTGCTGATGCTGTCAGCCACTGCGCCGCTCTGCT 900
Dh 1109 CTAGCTATGCACTTAATGATGTTGCTGATGCTGTCAGCCACTGCGCCGCTCTGCT 1168
Qy 901 CGAGATTAATCCCTCTCTCTCAGAAAGTTCCTCCGCGCAGTGGGATGTCCTGCTATA 960
Dh 1169 CGAGATTAATCCCTCTCTCTCAGAAAGTTCCTCCGCGCAGTGGGATGTCCTGCTATA 1228
Qy 961 AGCTCTTCTCTGATCAGAGGCTGCTGAGAGAGCCGCTCAAGAGATGAGCAGAG 1020
Dh 1229 AGCTCTTCTCTGATCAGAGGCTGCTGAGAGAGCCGCTCAAGAGATGAGCAGAG 1288
Qy 1021 C---CAGAGAGAACTGCTGTGACATTTGAAGATPAGAACGCGGAGAACTTTCAGCGA 1077
Dh 1289 CAATTGAGAAAGAAATTAATCTGTACGTTTGAAGATPAGAACGCGGAGAACTTTCAGCGA 1348
Qy 1078 GCGAGTGTGAGCTGAGAGAGCCGCTCAAGCGCTCTTGGAGAGAGAGCGAAAGAGAG 1137
Dh 1349 GCGAGTGTGAGCTGAGAGAGCCGCTCAAGCGCTCTTGGAGAGAGAGCGAAAGAGAG 1408

Qy 1138 GAGCGTTGGCTCAGCTGAGAGCGCGCCGAGCAGAGAGGAAAGCGGAGCGCCAGAG 1197
Dh 1409 GAGCGCTGGCCCGACCTGAGAGCGGCGGAGAGAGAGAGAGCGCTGGCCAGAG 1468
Qy 1198 CAGAGGCGAAAGCGGCGCTGAGAGCTGGAAGCAGCTGGAAGAGCAGGAGAGCTGAG 1257
Dh 1469 CAGAGGCGAAAGCGGCGCTGGAAGCAGCTGGAAGAGCAGCTGGAAGAGCAGGAGCTGGA 1528
Qy 1258 CCGAGCGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCGAAACGGGAA 1317
Dh 1529 CCGAG 1588
Qy 1318 CTGAAAGGCGAGCGCAACTTTGAATGGGAAACGGAACCGAGACAGAACTTCTGATCAG 1377
Dh 1589 CTTGAAAGGCGAGCGCAACTTTGAATGGGAAACGGAATCGAAGGAGAGAACTTCAATACAA 1648
Qy 1378 AGGAACAAAG 1437
Dh 1649 AGGAACAAAG 1708
Qy 1438 GAGTTAGAGCTGGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Dh 1709 GAGTTAGAGCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 1768
Qy 1498 TGTGACTGCGCAACCCAGAGCGAAAGAAATTTAGAGACGAACTTATGAGAGCTAAGA 1557
Dh 1769 TGTGACTGCGCAACCCAGAGCGAAAGAAATTTAGAGACGAACTTATGAGAGCTAAGA 1828
Qy 1558 ATTGCTGAATCAACCCAGCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Dh 1829 ATTGCTGAATCAACCCAGCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888
Qy 1618 CTTATTCAGAGAAACAGATACCTGATGACCAAGTTTAAACAGTCCAGCAGAAACAGTTTG 1677
Dh 1889 CTTATTCAGAGAAACAGATACCTGATGACCAAGTTTAAACAGTCCAGCAGAAACAGTTTG 1948
Qy 1678 CATAGAGCTGCTTTCTTAACTTCAAAAGAGCTTGAAGCAAAAGAGCTGGCCGCGAG 1737
Dh 1949 CATAGAGCTGCTTTCTTAACTTCAAAAGAGCTTGAAGCAAAAGAGCTGGCCGCGAG 2008
Qy 1738 CAGCTCCGAGAGAGCTGATT 1797
Dh 2009 CAGCTCCGAGAGAGCTGATT 2068
Qy 1798 GATGTTTCAACAACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
Dh 2069 GATGTTTCAACAACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128
Qy 1858 AAGCAGAGTCCCTGAGCTG 1917
Dh 2129 AAGCAGAGTCCCTGAGCTG 2188
Qy 1918 GAGTTAG 1977
Dh 2189 GAGTTAG 2248
Qy 1978 CTGAGAGATGTCAGCAGAGAG--GAGCAGCAGCGCCCGGAAACCCGAGAGAGAG 2034
Dh 2249 CTGAGAGATGTCAGCAG 2308
Qy 2035 AAGCTGAAG 2094
Dh 2309 AAGCTGAAG 2368
Qy 2095 ATGCAAG 2154
Dh 2369 ATGCAAG 2428
Qy 2155 CAGGACCTCTGCTTACACAG 2214
Dh 2429 CAGGACCTCTGCTTACACAG 2488

Db 1727 CTTATTCAGAAAAAGAGATCTCAATGACCAATTAAACAAAGTTGACGAAACAATTG 1786
 Qy 1678 CATAGAGACTCGGCTTCTTCACTCAAAAGAGCTTGGAGCAAGAGAGCTGGCCGCGAG 1737
 Db 1787 CACAGAGATTCATCTGTTACACTTAAAGAGCTTGAAGCAAAAGAACTAGTCCGCGAG 1846
 Qy 1738 CAGCTCCGGAGAGAGCTGAGCAGAGTGGAGAGAGAACAGGTCAAGCTGACAGAGATT 1797
 Db 1847 CACCTACGAGAACCACTGATGTAAGTGAAGAAAGAACTAGATCAAACTACAGAGATT 1906
 Qy 1798 GAGTTTTCACACCAAGCTGAAGAGAACTGAGAGATATCATGCAAGACAGCACTCCAG 1857
 Db 1907 GATATTTTCATATATCAGCTGAAAGAACTAGAGAAATACATATAGCAACAACCTCCAG 1966
 Qy 1858 AAGCAGAGTCCCTGAGCAGCGCCGACTGAACAGAAAAGAGCAGAGAGAGAAAGCTG 1917
 Db 1967 AACCAAAAGCTCATGAGGCTGAACGACTGAACAGAAAGAACAGAAAGAAAGATCAT 2026
 Qy 1918 GAGTTGAGAGAGAGAGAGAGAGCTCAGAGACGAGTTCAAGAAAGAGAGCAAGCAATGG 1977
 Db 2027 GAATTTGAAAAAAG 2086
 Qy 1978 CTGAGAGATGTCAG 2034
 Db 2087 CTGAGAGATGTCAG 2146
 Qy 2035 AGACTGAG 2094
 Db 2147 AAACGAG 2206
 Qy 2095 ATCAAG 2154
 Db 2207 GCAAG 2266
 Qy 2155 CAGGCACTCTGCTCTACAG 2214
 Db 2267 CAGGCACTCTGCTCTACAG 2326
 Qy 2215 AAGGTGATATTAACAG 2274
 Db 2327 AAGGTGATATTAACAG 2386
 Qy 2275 ATCCAGAGAGAGATATAGTCAT-----GATGATGAGAGAGAGAGAGAGAGAGAG 2319
 Db 2387 ATCCAGAGAGAGATATAGTCATGTTAAAGGAGATGAGTGAAGAGAGAGAGAGAGAGAG 2446
 Qy 2320 GAGCCAGAGATGCTTGGAG 2379
 Db 2447 GAACCCGCTGCTTGGAG 2506
 Qy 2380 GCAGAAAGATTCAGAAAGATGAGTTCCTCACTCCAGCCAAACAGTGAACGATCGACA 2439
 Db 2507 GCAGAGAAATCCAGAAATGAGTTCCTCCGCTCCAGTGAACAGTGAATGATTCACACA 2566
 Qy 2440 TCTGCCCCCTGCCCCAACTGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2499
 Db 2567 TCTGCCCCCTGCCCCAACTGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2626
 Qy 2500 TCTGAGCCCTCCAG 2559
 Db 2627 TCGAGACCTCCAG 2686
 Qy 2560 TCAAG 2619
 Db 2687 ACAGATGAG 2746
 Qy 2620 GTACCTAG 2679
 Db 2747 GTTCCAG 2806
 Qy 2680 TCTTCCCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2739
 Db 2807 TCTTCCCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2866

Qy 2740 CTGATCCCTGGAGAGCCAAAG 2799
 Db 2867 CTATATCTTTGAGAGAGCCAAAG 2926
 Qy 2800 ACCGTTTGGAG 2859
 Db 2927 ACCGTTTGGAG 2986
 Qy 2860 TTCCCAAGTCTTACGAG 2919
 Db 2987 TTCCCAAGTCTTACGAG 3046
 Qy 2920 ACTGAGCTTACGAG 2979
 Db 3047 TCTGAGTCTTACGAG 3106
 Qy 2980 GCAATCCCGAG 3039
 Db 3107 GTCGTTTGGAG 3166
 Qy 3040 TTAACTTTACAG 3099
 Db 3167 TTAACTTTACAG 3226
 Qy 3100 GGAACGTTGGAG 3159
 Db 3227 GGAACGTTGGAG 3286
 Qy 3160 GAGGAGCTCTGAG 3219
 Db 3287 GAGGAGCTCTGAG 3346
 Qy 3220 GTTATGCTTCTTACGAG 3279
 Db 3347 GTTATGCTTCTTACGAG 3406
 Qy 3280 ATTGATCCGAG 3339
 Db 3407 ATTGATCCGAG 3466
 Qy 3340 AAAAAAGCCAGATGAG 3399
 Db 3467 AAAAAAGCCAGATGAG 3526
 Qy 3400 AGCAAAATCAACCCCACTGAG 3459
 Db 3527 AGCAAAATCAACCCCACTGAG 3586
 Qy 3460 ATCCGAGATGAG 3519
 Db 3587 ATCCGAGATGAG 3646
 Qy 3520 ATCATCAACGCTCTCAACAG 3579
 Db 3647 ATCATCAACGCTCTCAACAG 3706
 Qy 3580 GTTGGGCTCTTCCATCAATTAATGTAAGCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3639
 Db 3707 GTTGGGCTCTTCCATCAATTAATGTAAGCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3766
 Qy 3640 TGA 3642
 Db 3767 TGA 3769

RESULT 7
 AF064244 AF064244 7247 bp mRNA linear PRI 21-NOV-1998
 LOCUS Homo sapiens Intersectin long form mRNA, complete cds.
 DEFINITION
 ACCESSION AF064244
 VERSION AF064244.1 GI:3859854
 KEYWORDS

SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 7247)
Antoniakis, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and Antonarakis, S.E.
TITLE Two isoforms of a human intersectin (ITSN) protein are produced by brain-specific alternative splicing in a stop codon
JOURNAL Genomics 53 (3), 369-376 (1998)
MEDLINE 99017974
PUBMED 9799604
REFERENCE 2 (bases 1 to 7247)
AUTHORS Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and Antonarakis, S.E.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue Michel-Servet, Geneva 4 CH-1211, Switzerland
FEATURES
source
1. 7247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..5272
/product="intersectin long form"
/protein_id="AAC78611.1"
/db_xref="GI:385985"
/translation="MAQEPPTFGGSLDIWATTVBERAKHQDFSLKPISGFITGDQA
RNFPOGSLPOPVLAIQIWMALADMDNDGMDVDFSIKAKLILKLGLOGLSAPVPM
KQOPVAISAPFGWGIASMPPLDPAVMDVDFSIKAKLILKLGLOGLSAPVPM
GAPVIOPLPAFPAHATLTPKSSFSRSGSOLNTKOKKASPDVAVSPVAMVAP
OSSRLYKROLPSHDKTMSGHLTGPOARTIMOSTLPAOALASIMNLDIDOKLTA
BEFILAMHIDVAMSGPULPYLPPEYIPSEFRVRSVSGISVLSSTSDVDRLEBPY
LEDEQOOLEKLPVTEDEKRENFERNLEKREKQALLIEQREKRELAQRELEBPY
RKEBEROERKQLEKREKREKREKREKREKREKREKREKREKREKREKREKREKRE
RNRROELNORKEQEDIVLAKAKKTEFELALNDKHOLEGLODIRLITTORE
ELESNKRRELR IAEITHLOOLOSQOOLGRLEPERKOLNDOLKOVONSLHDSIV
TLKRALKEKELAROLRDOLDEVEKETSKEIDIPNNOLKEIRELHINKOOLOKOS
MEAEKLOKEORKEITIELEKQREKQREKQREKQREKQREKQREKQREKQREKQRE
KREESVKKKDEKQ
KVVYRALYPPESRSHDEITLQPDIVWKGWVDSQGTGPMGLKGTGPMGLKGTGPM
NYAEKIPENVEVPAVPTDSTASAPKTLARETPAPLAPLAPLAPLAPLAPLAPLAPL
WPTSNKEPDTNDMAAOPSLTVPSAGOLROBSAFTPATATGSSPSPIVGOGEKVE
GLQOALYPMRAKKNHNLNFKNDVITTLBQDDMMFGEVGOQKMPFKSVYKLSGP
IRKSTSMDSGSESPASLKRVASPAKPVVSGEFTAMTYESSSQGLTVOQDVL
VTKDQDWMTGTVDGKAGFSPNRYRLKDSGSGTAGTGSIGKKPELAQVIASTYAT
GPEQLTLAPGOLILIRKKNPGWMEGELQARGKQKQKQKQKQKQKQKQKQKQKQKQ
EPPKSTALAAVCOVIGDYTAONDELAFNKGQIINVLNKEDPMKMGVNGVNGVNGV
PSNYVLTITDMDPSQWCSDLHLDMLPTREKRGQYTHELITVENVNDLOLVTEI
FOKPLMESELITKEVAMI FVNKELINCNILKILALVRRKMSGSKRPVMTGDIIS
AOLPMOPPIRFCRQLNGAALIQQKTEADPKEFVRLMDPCKMPLSSITLKP
MORVTRPPIIKNILENTPENHSHKHALKEAEELCSQVNEBKEKESDLEWI
QAHVCEGLSEQLVNSVTCNGPRKFLYAKKNNKILYGLFENDFILLQITK
PLGSSGTDKPSKSNLOYKWKPIPLNEVYLKPTPDGDEPIFHHSHLDYVTLR
AESINERTAMVQKTKAASELYIETKKRKRAYLVRSGRATGIGLMMVVEGELIKR
CSHGKSNPECVTVNGSOCHITKTIQDLPNPKNSNCFIRDSQEVLCITVERRO
FSPDPFLGRTEIRVADIKKQDSKGPVTKCLLHVPGEIVRLDLQLFDEP"
misc_feature
167..406
/note="encodes EH domain"
misc_feature
767..1936
/note="encodes EH domain"
misc_feature
2324..2524
/note="encodes SH3 domain"
misc_feature
2843..3019
/note="encodes SH3 domain"
misc_feature
3110..3286
/note="encodes SH3 domain"
misc_feature
3326..3520
/note="encodes SH3 domain"

Query Match 77.6%; Score 2826.8; DB 9; Length 7247;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 3164; Conservative 0; Mismatches 477; Indels 21; Gaps 3;
ORIGIN
BASE COUNT 2156 a 1652 c 1747 g 1692 t
misc_feature 3569..3748
/note="encodes SH3 domain"
misc_feature 3836..4390
/note="encodes GDS domain"
misc_feature 4649..4819
/note="encodes PH domain"
misc_feature 4895..5143
/note="encodes C2 domain"
Query 1 ATGGCTCAGTTCCACACCTTTGGTGTGTAGCTGTGATCTGGGCGCAATCTGTGAG 60
Db 107 ATGGCTCAGTTCCACACCTTTGGTGTGTAGCTGTGATCTGGGCGCAATCTGTGAG 166
Query 61 GAAAGGGCCAAAGCATGACACGACAGTTCCTTAGCCTGAAGCCGATAGCGGATTTTACT 120
Db 167 GAAAGGGCCAAAGCATGATGACAGTTCCTTAGTTAAAGCCAAATCTGTGATTCATTACT 226
Query 121 GGTGATCAAGGAGAACTTTTTCCTCATCTGGGTACTTCAAGCTGTCTTATGACACA 180
Db 227 GGTGATCAAGGATGAACTTTTTCCTCATCTGGGTACTTCAAGCTGTCTTATGACACAG 286
Query 181 ATATGGGGCTTAGCGGACATGAATTAAGATGAAGAGATGATCAAGTGAATTTTCATA 240
Db 287 ATATGGGCTAGTGTGACATGAATTAAGATGAAGAGATGATCAAGTGAATTTTCATA 346
Query 241 GCCATGAAGCTTATCAACCTGAAGCTTCAAGATATCAGTCCCTCCACACTTCCCTCCT 300
Db 347 GCTATGAAGCTTATCAACCTGAAGCTTCAAGATATCAGTCCCTCCACACTTCCCTCCT 406
Query 301 GTCATGAAGCAAGCAACGATGCTATTTCCAGTGCACACAGCATTTGGTATAGAGAGATT 360
Db 407 GTCATGAAGCAAGCAACGATGCTATTTCCAGGCAACACATTTGGTATAGAGAGATT 466
Query 361 GCTAGCATGCCACACCTCACAGCTGTGCTCTGTCGCAATGGGCTCATTCAGTTGTT 420
Db 467 GCCAGCATGCCACCGCTTACAGCTGTGCTCTGTCGCAATGGGCTCATTCAGTTGTT 526
Query 421 GAAATGTCGCCACCTTAGTATCTTCTGCTCCCTCAGACAGTCCCTCCCTGAGTAA 480
Db 527 GAAATGTCGCCACCTTAGTATCTTCTGCTCCCTCAGACAGTCCCTCCCTGAGTAA 586
Query 481 GGGGCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 587 GGGGCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 646
Query 541 AAGAGTCTCTCTCAGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 647 AAGAGTCTCTCTCAGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
Query 601 GCACATCATTCGATGTGCGCAGCGCCCTCCACAGACAGATAGGGCTGTGCTCAGTCA 660
Db 707 GCACATCATTCGATGTGCGCAGCGCGCTCCACAGACAGATAGGGCTGTGCTCAGTCA 766
Query 661 TCAAGGCTGAATACAGCAGTATTTCAACAGCCAGCAAAAATGATGAGTGAACCTTA 720
Db 767 TCAAGGCTGAATACAGCAGTATTTCAATATGATATCAAAATGATGAGTGAACCTTA 826
Query 721 ACAGGTCCTCCAGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 780
Db 827 ACAGGTCCTCCAGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 886
Query 781 TCAATATGGAATCTTTCTGACATTCATGATCAAGATGAGAAATCTCACTGAGAGATTTATC 840
Db 887 TCAATATGGAATCTTTCTGACATTCATGATGAGAGAAATCTCACTGAGAGATTTATC 946
Query 841 CTAGCTATGACCTTAATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

Db 947 CTGGCAATGCACTCATTTGATGTAGTATGTCTGGCCAAACCACTGCCACTGTCTCTCT 1006
Qy 901 CCGAATATACCTCCCT 960
Db 1007 CCGAATATACCTCCCT 1066
Qy 961 AGCTCTTCT 1020
Db 1067 AGCTCAACATCTGTATGATCAGAGGCTTCCAGAGAACCAATTTTAAAGATGAAACAACAA 1126
Qy 1021 C---CAGAGAAAGAACTGCTGTGACATTTGAGATTAAGAAAGGGAGAACTTCCAGCGA 1077
Db 1127 CAATTAATAAAGAAATTAATCTGTAACTTTGAAATTAAGAAAGCGAGAACTTTGAACGT 1186
Qy 1078 GCGAGTGTGAGCTGAGAAAGCCGCCCAAGCCCTCTTTGAGCAGCAGCGCAAGAGCAG 1137
Db 1187 GCGAACCCTGGAACTGGAGAAACGAGGCAAGCTCTCTCTGAAACAGCAGCGCAAGAGCAG 1246
Qy 1138 GAGCGTGTGCTCACTGAGCGCGCGGAGCAGAGAGAGAAAGCGGAGCGCGCAGAGAG 1197
Db 1247 GAGGCTGTGCGCCAGCTGTGAGCGGCGGAGCAGAGAGAGAGAGAGCGTGTAGCCCGCAGAG 1306
Qy 1198 CAGAGGCGCAGCGGCGCTGTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
Db 1307 CAGAGGCGCAG 1366
Qy 1258 CCGGAGCGAG 1317
Db 1367 CCGGAG 1426
Qy 1318 CTGAGAAAGGAG 1377
Db 1427 CTGAGAAAGGAG 1486
Qy 1378 AGAACAAG 1437
Db 1487 AGAACAAG 1546
Qy 1438 GAGTTAGAGCTCTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 1547 GAATTAAGAGCTCTTAAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
Qy 1498 TGTGAGCTGGGAG 1557
Db 1607 TGTGAGTTGACACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
Qy 1558 ATTGCTGAATCACCACTTAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 1667 ATTGCCGAATCACCACTTAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726
Qy 1618 CTTATTTCCAG 1677
Db 1727 CTTATTTCCAG 1786
Qy 1678 CATAGAGACTGCTTCTTAACCTCTCAAAAGAGCTTTGAGAGAGAGAGAGAGAGAGAG 1737
Db 1787 CACAGAGATTCTTGTGTACCTTAAGAGAGCTTTAGAGAGAGAGAGAGAGAGAGAGAG 1846
Qy 1738 CAGCTCCGGAG 1797
Db 1847 CACCTACGAG 1906
Qy 1798 GATGTTTTCACCAACAG 1857
Db 1907 GATGTTTTCACCAACAG 1966
Qy 1858 AAGCAGAGCTCCCTGAG 1917
Db 1967 AAGCAAAAGTCCATGAG 2026
Qy 1918 GAGTTAG 1977
Db 2027 GAATTAAG 2086

Qy 1978 CTGAGAGATGTGACAGAGAGAG---GAGCAGCAGCGCCCGGAGAAACCCACAGAGAGAGAC 2034
Db 2087 CTGGAGAGATGTGACAG 2146
Qy 2035 AGACTGAAG 2094
Db 2147 AAACGAG 2206
Qy 2095 ATGCAAG 2154
Db 2207 GCACAG 2266
Qy 2155 CAGGAG 2214
Db 2267 CAGGAG 2326
Qy 2215 AAAGGTGATATTAACGAG 2274
Db 2327 AAAGGTGATATTAACGAG 2386
Qy 2275 ATCCAGCAGAGAGATATAGTAT-----GGTGAATGAAGCCAGACTGGA 2319
Db 2387 ATCCAGCAGAGAGATATAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2446
Qy 2320 GAGCAGAGATGCTTGGAG 2379
Db 2447 GAACCCGAG 2506
Qy 2380 GCAG 2439
Db 2507 GCAG 2566
Qy 2440 TCTGCCCCCGCCCCCAACTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2499
Db 2567 TCTGCCCCCGCCCCCAACTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2626
Qy 2500 TCTGAGCCCTCCACAG 2559
Db 2627 TCGAGAGCCCTCCACAG 2686
Qy 2560 TCGAGAGCCCTCCACAG 2619
Db 2687 ACGAATGAG 2746
Qy 2620 GTACCTAGTGTGCTGAG 2679
Db 2747 GTTCAAGTGTGCTGAG 2806
Qy 2680 TCTTCCCATCTCCCGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2739
Db 2807 TCTTCCCATCTCCCGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2866
Qy 2740 CTGTATTCCTGTGAG 2799
Db 2867 CTATATTCCTGTGAG 2926
Qy 2800 ACCGTTCTGAG 2859
Db 2927 ACCGTTCTGAG 2986
Qy 2860 TTCCCAAGTCTTACGTGAG 2919
Db 2987 TTCCCAAGTCTTACGTGAG 3046
Qy 2920 ACTGAGCCCTACTGAG 2979
Db 3047 TCTGATTTCTTACGAG 3106
Qy 2980 GCCATTTCCGAG 3039
Db 3107 GTGTTTTGGGAG 3166

QY 3040 TTAACCTTTCAGCAGGAGATGATTTGTTACCAAGAAAAGTGTGACTGTGGACG 3099
 DB 3167 TTAACCTTTCAGCAGGAGATGATTTGTTACCAAGAAAAGTGTGACTGTGGACG 3226
 QY 3100 GGAACGGTGGGCGCAAGTCCGAGTCTTCCCTTCTAATATGTAGAGGTTAAAGATTCA 3159
 DB 3227 GGAACGGTGGGCGCAAGTCCGAGTCTTCCCTTCTAATATGTAGAGGTTAAAGATTCA 3286
 QY 3160 GAGGGCTTGGAACTGTGGGAAAACAGGAGATTTAGAAAAAACTGAATTGCCCCAG 3219
 DB 3287 GAGGGCTTGGAACTGTGGGAAAACAGGAGATTTAGAAAAAACTGAATTGCCCCAG 3346
 QY 3220 GTTATGCTTCTCAGCGCTGCTACTGTGTCCGAACTCACTGCTGCTGCTGCTGCTG 3279
 DB 3347 GTTATGCTTCTCAGCGCTGCTACTGTGTCCGAACTCACTGCTGCTGCTGCTGCTG 3406
 QY 3280 ATTCTGATCCGAAAAAGAACCCAGTGTGATGTTGGGAAAGGAACTTGAACCTGAGGG 3339
 DB 3407 ATTCTGATCCGAAAAAGAACCCAGTGTGATGTTGGGAAAGGAACTTGAACCTGAGGG 3466
 QY 3340 AAAAAGCGCAGATAGGAGTGTGTTTCCAGCAAAATTTATGTCAACTTAAAGCCCGGAGCA 3399
 DB 3467 AAAAAGCGCAGATAGGAGTGTGTTTCCAGCAAAATTTATGTCAACTTAAAGCCCGGAGCA 3526
 QY 3400 AGCAAAATCACTCCCACTAGCTTACCCAGAGCCGAGTCCAGCAGTGTGCTGCTGCTG 3459
 DB 3527 AGCAAAATCACTCCCACTAGCTTACCCAGAGCCGAGTCCAGCAGTGTGCTGCTGCTG 3586
 QY 3460 ATGGGATGATGATTAACATCCGCGCAGAGATGAGCACTAGCTTCAAGCAAAAGCCGAG 3519
 DB 3587 ATGGGATGATGATTAACATCCGCGCAGAGATGAGCACTAGCTTCAAGCAAAAGCCGAG 3646
 QY 3520 ATGATCAAGCTCTCTCAAGAGAGAGAGCCGAGTGTGAAAAGAGTGTGGGCAAA 3579
 DB 3647 ATGATCAAGCTCTCTCAAGAGAGAGAGCCGAGTGTGAAAAGAGTGTGGGCAAA 3706
 QY 3580 GTTGGGCTCTTCCATCAATTAATTAAGCTGACCAAGACATGGAACCCAGCCAGCAAA 3639
 DB 3707 GTTGGGCTCTTCCATCAATTAATTAAGCTGACCAAGACATGGAACCCAGCCAGCAAA 3766
 QY 3640 TG 3641
 DB 3767 TG 3768
 RESULT 8
 AF132672 3812 bp mRNA linear ROD 28-JUN-1999
 LOCUS Rattus norvegicus EH-domain/SH3-domain containing protein mRNA,
 DEFINITION complete cds.
 ACCESSION AF132672
 VERSION AF132672.1 GI:4838525
 KEYWORDS Rattus norvegicus.
 SOURCE Rattus norvegicus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 3812)
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T. C.
 TITLE BSH1/intersectin, a protein that contains EH and SH3 domains and
 binds to dynamin and SNAP-25. A protein connection between
 exocytosis and endocytosis?
 JOURNAL J. Biol. Chem. 274 (26), 18446-18454 (1999)
 MEDLINE 99303609
 PUBMED 10373452
 REFERENCE 2 (bases 1 to 3812)
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T. C.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAR-1999) Center for Basic Neuroscience and HHMI,
 Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX
 75235, USA
 FEATURES Location/Qualifiers

source 1. 3812
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 CDS 41..3481
 /note="BSH1"
 /codon_start=1
 /product="EH-domain/SH3-domain containing protein"
 /protein_id="A031026.1"
 /db_xref="GI:4838525"
 /translation="MAQPTPTFGSLDIWAIVTEBRKHHQDQFOSLKPISGITGDDA
 RNFPSGJLPOLVLAQIWLADNKKDQDVEFSIAMKLILKIQGLQVLPALPVM
 KOOPALISSAPFGIGMAGMPLTAVAVPMGSIIVMGSPLEVSVOALPVLPM
 GAPVVOPLPAPHPAATLPSKSSFSRSPGSLNKLKQASPVASPAEAAPV
 OSSRLKRLQFNHSDKTMGHLTPQPARITLWQSSLPQQLASINLUSITDDGKTA
 EEFILAMHLIDVAMSQPLPVLPIPPSFRRVVGSGMSVSSSADQRPKPEBCKLR
 SEDQOQVEKRLPVTFEDKRENERNGNLEKROALLEQOREDELRLOLEWER
 KEREROERKROLELEKOLEERLEERREERKEIRREAREKELERLOLEWER
 NROELLTNRNKOEGIVLKKARKTLFEFLNKKHOLEKLDKIDLCRIATOROE
 IESTNRSRLRLAETHLOOQLOESQOQIIVKQIISDOLKVOQOQSLHRSLLT
 LKRLAKELARQQLRELDVEKERTSKLOEIDVNNQKLEIRHISKOQLOKORSI
 EAERLQKEQERKLELEKQERKQRRVQERQWQEHVQOEQRPRKPEBCKLR
 EEDVKKKEAEERAKPEVODKQSLFPHQEPAPQAPMPTTEKGLTISAGESAVV
 YRALVPFESRSHDEITLIOGDIIVMKGEVDESOTGEGMLGEGPKGTGWFPAVYA
 EKIPENEIPTPAKPVTDITSAAPKXLALETTPALPVTSSSEPTTNMADFSSTWS
 STNEKRETDWMDTMAQPSLTVPASQALORSAFPTPATAGSSPSVLOGEKVEGLQ
 AQLVTPRAKNDNHLNPNKSDVITVLBQDDMFVGGKGFPSYKYLISGPRK
 STSIDGTPAPSLRSLKRVASPAKPAIPGEILAQVLAISTATGEBUTLAPGLILR
 KKNPGWMBEGLQARGRKQIIGWFPANVYVLSGPSKIPTELPTAVQPAVCOVIG
 MYDTAQNDDELAFSKQIINVLSKEDPDPMKGEVSGVGLPFSNVVXKLTTMDPSEQ
 "

BASE COUNT 1088 a 972 c 1058 g 694 t
 ORIGIN
 Query Match 77.3%; Score 2816.2; DB 10; Length 3812;
 Best Local Similarity 87.7%; Pred. No. 0;
 Matches 3208; Conservative 0; Mismatches 218; Indels 231; Gaps 3;
 QY 1 ATGGCTGATTTCCCAACCTTTGGTGTAGCTGTGATGTTCTGGCCATACCTGTGAG 60
 DB 41 ATGGCTGATTTCCCAACCTTTGGTGTAGCTGTGATGTTCTGGCCATACCTGTGAG 100
 QY 61 GAAAGGCGCAAGCATGACAGACAGTCTTCCCTGAGAGCCGATAGGGGATTTATTACT 120
 DB 101 GAAAGGCGCAAGCATGACAGACAGTCTTCCCTGAGAGCCGATAGGGGATTTATTACT 160
 QY 121 GGTGATCAAGCAGAACTTTTTCATCTGGGTTACCTCAGCTGTCTTACGACAA 180
 DB 161 GGTGATCAAGCAGAACTTTTTCATCTGGGTTACCTCAGCTGTCTTACGACAA 220
 QY 181 ATATGGGCGCTAGCGAGCATGATTAACGATGAAGATGATCAAGTGAATTTTCCATA 240
 DB 221 ATATGGGCGCTAGCGAGCATGATTAACGATGAAGATGATCAAGTGAATTTTCCATA 280
 QY 241 GCCATGAAGCTTATTAACCTGAAGTACAAAGATTAACGCTCCCTCCACACTTCCCTCT 300
 DB 281 GCCATGAAGCTTATTAACCTGAAGTACAAAGATTAACGCTCCCTCCACACTTCCCTCT 340
 QY 301 GTCATGAACAGCAACAGTGTCTTTCAGTGACAGCAGATTTGTATAGAGGAGATT 360
 DB 341 GTCATGAACAGCAACAGTGTCTTTCAGTGACAGCAGATTTGTATAGAGGAGATT 400
 QY 361 GCTAGCAGTCCACCACTCAACAGCTGTCTCTGTGCAATAGGCTTCATTCAGTTT 420
 DB 401 GCTAGCAGTCCACCACTCAACAGCTGTCTCTGTGCAATAGGCTTCATTCAGTTT 460
 QY 421 GGAATGCTCCACCTTGTATCTTCTGCTCCAGAGAGAGTCCCTCCCTGAGCTTAC 480
 DB 461 GGAATGCTCCACCTTGTATCTTCTGCTCCAGAGAGAGTCCCTCCCTGAGCTTAC 520
 QY 481 GGGGCTCTCCCGTATACAGCTGTGCTGCTGTTGGCATCTTCAGCAGCACAATGGCA 540
 DB 521 GGGGCTCTCCCGTATACAGCTGTGCTGCTGTTGGCATCTTCAGCAGCACAATGGCA 580

QY 541 AAGATTCTTCTTCAGCAGATCTGGTCCAGGGTCACAATTAACTAAGTTACGANG 600
 Db 581 AAGAGTTCTTCTTCAGCAGATCTGGTCCAGGGTCACAATTAACTAAGTTACGANG 640
 QY 601 GCACCAATCATTCATGTCGCGCAGCGCCCTCCAGCAGCAAGATGGCTGTGCTCAGTCA 660
 Db 641 GCACCAATCATTCATGTCGCGCAGCGCCCTCCAGCAGCAAGATGGCTGTGCTCAGTCA 700
 QY 661 TCAGAGCTGAATATACAGGCGATTATTCACAGCCAGCAAACTATGATGAGCACTTA 720
 Db 701 TCAGAGCTGAATATACAGGCGATTATTCACAGCCAGCAAACTATGATGAGCACTTA 760
 QY 721 ACAGATCCCGAGGCAAGAACTATTCATGCATCAAGTTATCCCAAGCTCAGCTGGCT 780
 Db 761 ACAGATCCCGAGGCAAGAACTATTCATGCATCAAGTTATCCCAAGCTCAGCTGGCT 820
 QY 781 TCAATATGGAATCTTTCTGACATTGATCAAGATGAAAACTCATGACAGAAATTTATC 840
 Db 821 TCAATCTGGAATCTTTCCGACATTGATCAAGATGAAAACTCATGACAGAAATTTATC 880
 QY 841 CTAGCTATGCACTTAATGATGTTGTCATGTCTGGTCAAGCACTGCCGCCCTCTGCT 900
 Db 881 CTAGCTATGCACTTAATGATGTTGTCATGTCTGGTCAAGCACTGCCGCCCTCTGCT 940
 QY 901 CCAGATATCATCCCTCCTCTTCAGAAAGATTGCTCCGCGAGTGGAGATGTCCTGATA 960
 Db 941 CCAGATATCATCCCTCCTCTTCAGAAAGATTGCTCCGCGAGTGGAGATGTCCTGATA 1000
 QY 961 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGCTGTCAGAGATGAGCAGCAG 1020
 Db 1001 AGCTCTTCTGCTCAGAACCCAGCGGCTGCCGAGAGCCATCTGCAAGATGAGCAGCAG 1060
 QY 1021 CCAGAGAAAGAACTGCTGTGATCAATTTGAAAGATTAAGACGGAGAACTTTCAGACGAG 1080
 Db 1061 GTGGAAGAAAGCTGCTGTGATCAATTTGAAAGATTAAGACGGAGAACTTTCAGACGAG 1120
 QY 1081 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGCAGCAGCGCAAGAGCAGAG 1140
 Db 1121 AACTGTGAGCTGAGAGAGCGCGCAAGCGCCCTCTTGAGCAGCAGCGCAAGAGAGCAGAG 1180
 QY 1141 CGGTTGGCTCAGCTGAGAGCGCGCCGAGCAGAGAGAGAAAGCGGGAGCGCCAGAGCAG 1200
 Db 1181 CGGCTGGCTCAGCTGAGAGCGCGCCGAGCAGAGAGAGAAAGCGGGAGCGCCAGAGCAG 1240
 QY 1201 GAGGCGAAGCGGAGCTGAGAGCTGAGAGAGCTGAGAGAGCGGAGCGGAGCTGAGAGCG 1260
 Db 1241 GAGGCGAAGAGAGCTGAGAGAGCTGAGAGAGCGGAGCGGAGCGGAGCTGAGAGCG 1300
 QY 1261 CAGCGAGAGAGAGAGAGAGAGATGAGAGGCGCGAGCGCCGCAAAACGGGAACTG 1320
 Db 1301 CAGAGAGAGAGAGAGAGAGAGATGAGAGGCGCGAGCGCCGCAAAACGGGAGCTC 1360
 QY 1321 GAAAGCAGAGAGCAACTTAATGAGAACGGAACCGGAGACAGAAATCTCTGAATCAGAG 1380
 Db 1361 GAAAGCAGAGAGCAACTTAATGAGAACGGAACCGGAGACAGAAATCTCTGAATCAGAG 1420
 QY 1381 AACCAAG 1440
 Db 1421 AACCAAG 1480
 QY 1441 TTGAGAGCTCTGATGATCAAAAAAGCATCAGCTGAGAGAGAAAACTTCAGATATCAGGT 1500
 Db 1481 TTGAGAGCTCTGATGATCAAAAAAGCATCAGCTGAGAGAGAAAACTTCAGATATCAGGT 1540
 QY 1501 CGAGCTGGCAACCCAGAGGCAAGAAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 Db 1541 CGAGCTGGCAACCCAGAGGCAAGAAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
 QY 1561 GGTGAATATCACCCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 Db 1601 GGTGAATATCACCCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1660
 QY 1621 ATTCCAGAGAAACAGATCTCAGTGAACGATTAAAAAGTCCAGAGAGAGAGAGAG 1680

Db 1661 ATTCCAGAGAAACAGATCTCAGTGAACGATTAAAAAGTCCAGAGAGAGAGAGAGAGAG 1720
 QY 1681 AGAGATCTGCTTTCTTACCTTCAAAAAGAGCTTGGAGCAAAAGAGAGAGAGAGAGAG 1740
 Db 1721 AGAGATCTGCTTTCTTACCTTCAAAAAGAGCTTGGAGCAAAAGAGAGAGAGAGAGAG 1780
 QY 1741 CTCCGGAGAGAGCTGAG 1800
 Db 1781 CTCCGGAGAGAGCTGAG 1840
 QY 1801 GTTTTCAACCAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 Db 1841 GTTTTCAACCAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1900
 QY 1861 CAGAGGTCCTGAG 1920
 Db 1901 CAGAGGTCCTGAG 1960
 QY 1921 TTGAG 1980
 Db 1961 TTGAG 2020
 QY 1981 GAGCATGTGAG 2040
 Db 2021 GAGCATGTGAG 2080
 QY 2041 AAGAGGAG 2100
 Db 2081 AAGAGGAG 2140
 QY 2101 GAG 2160
 Db 2141 GAG 2197
 QY 2161 CCTGTGCTCAACAG 2220
 Db 2198 CCTGTGCTCAACAG 2257
 QY 2221 GTATATTAACGAG 2280
 Db 2258 GTATATTAACGAG 2317
 QY 2281 CCAGAGATATATGATCAT-----GATGATGAGAGAGAGAGAGAGAGAGAGAGAG 2325
 Db 2318 CCAGAGATATATGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2377
 QY 2326 GATGAGCTTGGAG 2385
 Db 2378 GATGAGCTTGGAG 2437
 QY 2386 AAGATTCAG 2445
 Db 2438 AAGATTCAG 2497
 QY 2446 CTTGCCCCCAAACTGGCTCTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2505
 Db 2498 CTTGCCCCCAAACTGGCTCTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2557
 QY 2506 CCTTCCAAAACCCCAAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2565
 Db 2558 CCTTCCAAAACCCCAAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2617
 QY 2566 GAG 2625
 Db 2618 GAG 2677
 QY 2626 AGTGTGCTGAG 2685
 Db 2678 AGTGTGCTGAG 2737
 QY 2686 CCAATCTCCCTCTGAG 2745

```

Db 2738 CCTTCTCCCTGCTGGGCGAGGAGTGAAGAGGCTCCAGACCAAGCCCTGTAT 2797
QY 2746 CCTTGGAGAGGCAAAAACCAACCACTTAATTTTAAACAAAGAGGTATCAACCGTT 2805
Db 2798 CCTTGGAGAGGCAAAAACCAACCACTTAATTTTAAACAAAGAGGTATCAACCGTT 2857
QY 2806 CTGGAACACAGACACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2865
Db 2858 CTGGAACACAGACACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2917
QY 2866 AAGCTTACGTAAGTAATCTATTTTACGAGGCGCGTGAAGAAATCCACAGATGATATGTC 2925
Db 2918 AAGCTTACGTAAGTAATCTATTTTACGAGGCGCGTGAAGAAATCCACAGATGATATGTC 2977
QY 2926 CCTTACGTAAGTAATCTATTTTACGAGGCGCGTGAAGAAATCCACAGATGATATGTC 2985
Db 2978 CCTTACGTAAGTAATCTATTTTACGAGGCGCGTGAAGAAATCCACAGATGATATGTC 3037
QY 2986 CCGGAGAGAGGTTTATTTCCATGTACATACAGAGATTTTACAGCAAGATTTTAAAC 3045
Db 3038 CCGGAGAGGTTTATTTCCATGTACATACAGAGATTTTACAGCAAGATTTTAAAC 3046
QY 3046 TTTTACAGCAAGAGGATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3105
Db 3047 ----- 3046
QY 3106 GTGGGCGACAGATCGGAGTCTTCCCTTCTACTATGTAGAGCTTAAAGATTCAAGAGGC 3165
Db 3047 ----- 3046
QY 3166 TCTGGAAGTGTGGGAAACAGGAGATTAGAGAAAAAAGCTGAAATTTGCCAGTTATT 3225
Db 3047 ----- 3064
QY 3226 GCTTCTTACGCTGCTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3285
Db 3065 GCTTCTTACGCTGCTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3124
QY 3286 ATCCGGAAGAAACAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3345
Db 3125 ATCCGGAAGAAACAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3184
QY 3346 CCGCAGATGAGGTTTCTTCCAGCAAAATTTATGTCAAACTTTAGCCCGGAAACAGCAAA 3405
Db 3185 CCGCAGATGAGGTTTCTTCCAGCAAAATTTATGTCAAACTTTAGCCCGGAAACAGCAAA 3244
QY 3406 ATACCCCAACTAGAGTACCAAGACGCGACGAGTGTGTGTGTGTGTGTGTGTGTGTGT 3465
Db 3245 ATACCCCAACTAGAGTACCAAGACGCGACGAGTGTGTGTGTGTGTGTGTGTGTGTGT 3304
QY 3466 ATGTACGATTACACCGCGCAGAAAGATGACGAACTTACGCAAAAGGCCAGATCATC 3525
Db 3305 ATGTACGATTACACCGCGCAGAAAGATGACGAACTTACGCAAAAGGCCAGATCATC 3364
QY 3526 AAGCTCTTCAACAGAGGAGACCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3585
Db 3365 AAGCTCTTCAACAGAGGAGACCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3424
QY 3586 CTCTTCCATCCATTTTGTAAAGCTGACACAGACATGAGACCCAGCCAGCAATGA 3642
Db 3425 CTCTTCCATCCATTTTGTAAAGCTGACACAGACATGAGACCCAGCCAGCAATGA 3481

```

```

RESULT 9
AF032118 4103 bp mRNA linear VRT 13-NOV-1998
LOCUS Xenopus laevis interseclin mRNA, complete cds.
DEFINITION
ACCESSION AF032118
VERSION AF032118.1 GI:2642624
KEYWORDS
SOURCE Xenopus laevis.
ORGANISM Xenopus laevis.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
1 (bases 1 to 4103)
AUTHORS Yamabhai, M., Hoffman, N.G., Hardison, N.L., McPherson, P.S.,
Castagnoli, L., Cesareni, G. and Kay, B.K.
TITLE Interseclin, a novel adaptor protein with two Eps15 homology and
five Src homology 3 domains
JOURNAL J. Biol. Chem. 273 (47), 31401-31407 (1998)
MEDLINE 99030416
PUBMED 9813051
REFERENCE
2 (bases 1 to 4103)
AUTHORS Hoffman, N.G., Hardison, N.L., Yamabhai, M. and Kay, B.K.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1997) Pharmacology, University of Wisconsin, 1300
University Ave, Madison, WI 53706-1532, USA
FEATURES
source
location/Qualifiers
1..4103
/codon_start=1
/function="involved in endocytosis"
/note="EH domain and SH3 domain containing protein;
similar to Intersectin binding proteins Ibp1, encoded by
GenBank Accession Number AF057285, and Ibp2, encoded by
GenBank Accession Number AF057286, and mouse
Rab/Rip, encoded by GenBank Accession Number AF057287"
/codon_start=1
/product="Intersectin"
/protein_id="AAC73068.1"
/db_xref="GI:2642625"
/translation="MAQFGTFFGNDIWAITVEERAKHDQFHLKPTAGYITGDQ
RNFELSGLPQPLAOLMALADNNDNRDGLDFSLAMKLIKLOGYPLSLPDM
LKQPVMAAIVAGFGMSGIVGIPPLAIVAPVMSIPVGMSPPLVSSVPPVPLSM
GAPAVIOSHPAFASATLPEKSSFFGSAVSGSINTLQAGSDVAPPLVEMAPPS
SSRLKROLFNSODKTMGNTGPQARTLMOSSLPOSQATTWISDITDQKLTAE
EFLIAMLIDVAMSGOPLPILPEYIPSPFRVRSGLSISVSVDQRLPEEPE
EEPONADKRLPTEFKRENFPERGNLIEKREQLLEOQKOEKRLAOLERAEOR
ERRODOERKROODLEKORELEKREERERKEIEREAKRELEKROLEMER
RROELNORREBEDIVYLKAKKTELELEALNDKKHLEGLDORLITRHEI
ESTNKSRELIETITLHQOOLQESQOLGKMTPEKSLIDOLKOVQNSLHPSLTL
KRALETKEIGROQLDQLEVEREKTRAKQIEDVFNNOLEKELYNKQOPQODPE
TEKIKOKELEKRTSELDKLKEEDKRMLEBDKLMORVQOEERLYNFODEKEREES
VOCKEVEKPEIOEKPKPHOPPEPKLGGQIIPMNTTERAPLTINGQVKKVYRAL
YPPDARSHDEITTEPGDILIMVDSOSGEBMGLGELKGTGMPANVYAEEMSPSPS
TTTPAEETTKPTVHVAPSPVPAALFTNTSTNNMADSSSTPTNNTKVESDMNDT
WAQPSLTVPASQOHRQSAFTPATVYGSPPSVLQGEKVBGLQALYPRAKDN
HLNFKNKDVITVLEODMMWFGEVOQKGFPSYVKLISGPLRKSTSIDTSSES
SLKRVSSPAFKPAIOGEEYISMYTESNEGDLTFQOGLIIVIKKDGMMGTVEK
TGVPSPSYVAPKDESEAGSGGSGTSGKKEIIOVILASVATAPBQULTAPGOLILIR
KNPDGMBEGLDARGGKROIGWPPANVYVYLISRGNTKSTPTEPPTSLPTECOYIG
MYDIAONDELAFSGKQIVINLVNKEDPDMKELNGHVLPSNVYKLTTPMDPSQO
FRIGVKKPAGGIPATGDRPFTLPPFRDGPSSLPAFAQPLSVVMIRFCTARFCPD
MNVYKINI"
misc_feature
253..498
/note="encodes Eha domain"
misc_feature
850..1101
/note="encodes Ehb domain"
misc_feature
2386..2571
/note="encodes SH3A domain"
misc_feature
2881..3057
/note="encodes SH3B domain"
misc_feature
3148..3324
/note="encodes SH3C domain"
misc_feature
3364..3558
/note="encodes SH3D domain"
misc_feature
3607..3786
/note="encodes SH3E domain"
BASE COUNT 1290 a 920 c 1029 g 856 t 8 others
ORIGIN
Query Match 55.7%; Score 2027.6; DB 5; Length 4103;
Best Local Similarity 73.6%; Pred. No. 0;

```

Matches 2685; Conservative 4; Mismatches 918; Indels 39; Gaps 7;

QY 1 ATGGCTAGTTTCCACACCTTTGCGTGTAGCTGTGATGTCTGGCCATTAACCTGTGAG 60
 Db 193 ATGGCTAGTTTGAAGCTCCGTTTGGGGGTAATTTGGAATCTGGGCAATTAACGGTGGAG 252

QY 61 GAAAGGGCCAGAGATACCAAGAGCTTCTTAGCTGAAGCCCAATACGGGATTTTAACT 120
 Db 253 GAAAGAGCTTAACATGACAGAGCTTCAATGAGGATGAGCAACATGAGTTCTCCATA 312

QY 121 GGTATCAAGCAGAGAACTTTTTCATCTGGGTACCTAGCTCTCTTAGCACAA 180
 Db 313 GGTATCAAGCTAGAAATTTTTCCTTCACTCGGCTCCCAAGCTGTGTGGACAG 372

QY 181 ATATGGGCGGTAGCGACATGAATTAACGATGAGAGATGATTAAGTGAATTTTCAAT 240
 Db 373 ATATGGGCGGTAGCGACATGAATTAACGATGAGAGATGAGCAACATGAGTTCTCCATA 432

QY 241 GCGATGAGCTTATCAAACTGAAGCTACAAAGATATCAGTCCCTCCACACTTCCCTC 300
 Db 433 GCTATGAATTAATCAAGCTAAACATCAAGTTACCCGTGCTTATTTCTCCCTTCC 492

QY 301 GTCTAAGAACAGAACCAAGTGTCTAT-----TTCAATGACACAGCAATTTGTATAGA 354
 Db 493 AATATGTGAGAACACCGGTGCTGCTGCTGTGCTGTGCGGGGTTCGGGATGAGT 552

QY 355 GGGATTTGCTAGCATGCGACACACTCAGAGCTGTGCTCTGTGCAATGGGCTCCATTCCA 414
 Db 553 GGAATGTGGGCACTTCTCTCCCTAGAGCTGTAGCCCTGTGCAATGGCAATCATCCCA 612

QY 415 GTTGTGGAATGTCTCAACCCTTAGTATCTTGTCTCTCCAGCAGCACTGCTCCCTG 474
 Db 613 GTATGAGAAATGTCTCTCTCTGTGTCTGTCTGTCTCTCA-----CAGTTCACCTTG 666

QY 475 GCTAAGGGGCTCTCCCGCTACACAGCTCTGCTGCTGTGCGATTCCTGAGCGACA 534
 Db 667 TCTAAGGGGCTCTCCAGTAATCAAGTCTCATCTGTGCTTGTGCTCACTCT---GCCACA 723

QY 535 TGCCCAAGAGTCTTCTCTCAGCAGATCTGTGCTCAGGCTCACAATTAACATAAGTTA 594
 Db 724 TTGCCAAGAGTCTTCTCTCAGCAGATCTGTGCTCAGGCTCACAATTAACATAAGTTA 783

QY 595 CAGAGGCAATCATTCATGTGCTGCGAGGCTCTCCAGCAGCAGATGAGCTGTGCT 654
 Db 784 CAGAAAGCTCAGTCTTCAATGTCTGAGCCCTCCCACTTGTGTAATGGGCTGTGCCA 843

QY 655 CAGTCAAGGCTGAATTAACAGCAGTTATTCACAGCAGCAGCAAACTATGATGGA 714
 Db 844 TCGTCTTCAAGATTGAATGACAGCAATTTGTTACAGCAGCAGCAAACTATGATGGA 903

QY 715 CACTTAACAGGCTCCAGGCAAGAACTATTTCTATGCAATCAAGTTTACCCAGGCTCAG 774
 Db 904 AATTTAAGGCTCTCAGGCAAGAACTATTTCTATGCAATCAAGTTTACCCAGGCTCAG 963

QY 775 CTGGCTTCAATGGAATCTTCTGACATTTGATCAAGATGGAATTCCTCAAGAGAA 834
 Db 964 CTAGCTAATATGGAATCTTCTGACATTTGATCAAGATGGAATTCCTCAAGAGAA 1023

QY 835 TTTATCTGATGATGACCTAATTTGATGTTGACATGCTGTGAGCAGCACTGCGCCGT 894
 Db 1024 TTTATTTGATGATGACCTAATTTGATGTTGACATGCTGTGAGCAGCACTTCTCCAAATC 1083

QY 895 CTGCTTCCAGAAATCAATCCCTCTCTCTTCAAGAGATTCCTCCGAGTGGGATGTC 954
 Db 1084 CTGCTTCCAGAAATCAATCCCTCTCTCTTCAAGAGATTCCTCCGAGTGGGATGTC 1143

QY 955 GTCATTAAGCTCTTCTGTGATGAGGCTGCTGAGAGCGGTGTCAGAGGATGAG 1014
 Db 1144 ATTAATAGCTCTGTGATGAGGCTGCTGAGAGCGGTGTCAGAGGATGAG 1203

QY 1015 CAGCAGCAGAGAGAACTGCTGTGACATTTGAAGATTAAGAGCGGAGAACTTCCAG 1074
 Db 1204 CAAAGCGGAGAGAAAGCTCCAGTGAATTTGAAGACAAAGAGGAGAACTTTGAG 1263

QY 1075 CGAGGCACTGTGAGCTTGAAGAAAGCCGCCCAAGCCCTCTTGAAGACGACGCCCAAGAG 1134
 Db 1264 CGAGGAAACCTAGAAATTAAGAAAGAGAGACAGGCACTTTTAAGAGAGCGCCCAAGAG 1323

QY 1135 CAGAGAGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGCGAGCCCGAG 1194
 Db 1324 CAGAGAGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAGAAAGCGAGCGCGCA 1383

QY 1195 GAGCAGAGAGCCAAAGCGGAGCTGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTG 1254
 Db 1384 GATCAAGAGCGGAGAGCGGAGCAAGACCTGAGAAAGACAGCTGAGAGAAACAGAGAACTG 1443

QY 1255 GAGGCGAGCGAGAGAGAGAGAGAGAGAGAGATGAGAGGCGGAGCGCGCAAAAGCG 1314
 Db 1444 GAGGCGAGCGGAGAGAGAGAGAGAGAGAGAGAGATGAGAGGCGGAGCGCGCAAAAGCG 1503

QY 1315 GAACCTGAAAGGCGAGCAGCACTTGAATGGAAACGGAACCGAGACAGAACTCTGAAT 1374
 Db 1504 GAACCTGAAAGGCGAGCAGCACTTGAATGGAAACGGAACCGAGAGCAAGAACTCTGAAT 1563

QY 1375 CAGAGGAAACAG 1434
 Db 1564 CAAAGGAATAG 1623

QY 1435 TTTGAGTTAAGGCTGGAATGACAAAAGACATCAGCTAGAGAAAGAACTTCAAGATATC 1494
 Db 1624 TTTGAGTTAAGGCTGGAATGACAAAAGACATCAGCTAGAGAAAGAACTTCAAGATATC 1683

QY 1495 AGGTGCACTGCGCAACCGAGGCAAGAAATTAAGAGACAGAACTTGAAGAGCTA 1554
 Db 1684 CGGTGCGCTCTCACTCAACCGCAAGAAATTAAGAGACAGAACTTGAAGAGAGCTA 1743

QY 1555 AGAATTCGAAATCAACCCCTTAACAGACAGTTCAGAGAACTCAGCAAAATGCTTGA 1614
 Db 1744 AGAATTCGAAATCAACCCCTTAACAGACAGTTCAGAGAACTCAGCAAAATGCTTGA 1803

QY 1615 AGACTTATTCAGAGAAACAGATCACTCAGTGAACAGTTTAAACAGTCCAGCAGAACT 1674
 Db 1804 AATATGATTCCTGAGAAACAGTCCCTTATGATCAATTAAGCAGGTTCAACAAACAGT 1863

QY 1675 TTGATATGAGAGTCTCTTCTTCAACAAAGAGCTTGAAGCAAAAGAGTGGCCCG 1734
 Db 1864 TTGATATGAGAGTCTCTTCTTCAACAAAGAGAGCTTGAAGCAAAAGAGTGGAGCC 1923

QY 1735 CAGCAGCTCCGAGAGAGCTGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
 Db 1924 CAGCAGCTCCGAGAGAGCTGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1983

QY 1795 ATTGATTTTTCAACCAACAGCTGAAGAGAACTGAGAGAGATACATGCAACAGCACTC 1854
 Db 1984 ATCAGCTATTTTAATTAACCAATTAAGAGAACTGAGAGAACTGAGCAACAGCAGTTC 2043

QY 1855 CAGAGAGAGAGTCCCTGAGGAGAGCGGAGCTGAGCAGAGAGAGAGAGAGAGAGAGAG 1914
 Db 2044 CAGAGAGAGAGTCCCTGAGGAGAGCGGAGCTGAGCAGAGAGAGAGAGAGAGAGAGAG 2103

QY 1915 CTGAGATTGAAG 1974
 Db 2104 AGCAGCTGAGATTAAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2163

QY 1975 TGGCTGAGAGATGTGAG 2034
 Db 2164 TGGCTGAGAGATGTGAG 2217

QY 2035 AGACTGAAG 2094
 Db 2218 AAG 2271

QY 2095 ATGCAAGACAG 2154
 Db 2272 ATCAG 2331

QY 2155 CAGGCACTCTGCTCTACCAAGAGAAAGCCGCTTACCTTTCTGACAGAGAGTGA 2214
 Db 2332 CAGATTCCTGGATGAATACGAGAAAGCCCTCTAACGAT---CAACGAGGGCGATGG 2388
 QY 2215 AAGGTGTATTTATTCAGGACCGCTGTATCCCTTTTGAATCCAGAAATCAGATAGATTCACC 2274
 Db 2389 AAGGTGTATTTATTCAGGACCGCTGTATCCCTTTTGAATCCAGAAATCAGATAGATTCACC 2448
 QY 2275 ATCCAGACGAGAGATATATATCTATGTGTATGAAGAGCAGATCTGAGAGCAGAGATGGCTT 2334
 Db 2449 ATTTGAGCTCTGGKACATTTATCATGTGTATGAAGAGCAGAACTGGAGAGCTGGGTGGCTG 2508
 QY 2335 GAGAGAGCTGTAAG 2394
 Db 2509 GAG 2568
 QY 2395 GAAATAGAGTTTCCCTCAGCAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
 Db 2569 GAG 2628
 QY 2455 AACTGCTCTGCTGAG 2514
 Db 2629 CATGTGGAC-----CTTCCCTGTTGACCTGACCTTTTACAAATACATCTACA 2679
 QY 2515 ACCCCCAACAGCTGGGAG 2574
 Db 2680 AATTCCAACTGAGGAG 2739
 QY 2575 GAAAGGAG 2634
 Db 2740 GAG 2799
 QY 2635 CAGTTACGAG 2694
 Db 2800 CAG 2859
 QY 2695 GTCTCTGGGAG 2754
 Db 2860 GTACTGTGTGAG 2919
 QY 2755 GCCAAAAAG 2814
 Db 2920 GCAAAAG 2979
 QY 2815 CAG 2874
 Db 2980 CAG 3039
 QY 2875 GTGAACTCATTTTCAAGGAG 2934
 Db 3040 GTAAAGCTTATATCCGATCAGTAAG 3099
 QY 2935 AGTCTGTCTATGTAAAG 2994
 Db 3100 AGCCAGCGAG 3159
 QY 2995 GAGTTATTTCCATGTACATACAG 3054
 Db 3160 GAATACATATCCATGTACATATGAG 3219
 QY 3055 GGGAG 3114
 Db 3220 GAG 3279
 QY 3115 AAGTCCGAG 3174
 Db 3280 AAG 3339
 QY 3175 GCTGGAG 3234
 Db 3340 GAG 3399
 QY 3235 GCTGCTACTGTCTCCGAG 3294

Db 3400 GCAGCACTCTCTTGAAG 3459
 QY 3295 AAG 3354
 Db 3460 AAG 3519
 QY 3355 GGGTGTCTTCCAG 3414
 Db 3520 GGATGTTTCCAG 3579
 QY 3415 ACTGAGTACCCAG 3474
 Db 3580 ACKGAG 3639
 QY 3475 TACAG 3534
 Db 3640 TATATCCGAG 3699
 QY 3535 AACAG 3594
 Db 3700 AACAG 3759
 QY 3595 TCCAAATATGTAAAGCTGACAG 3640
 Db 3760 TCCAACTAAG 3805

RESULT 10
 AK027846
 LOCUS
 DEFINITION
 Homo sapiens cDNA FLJ14940 fis, clone PLACE1010942, highly similar
 to Homo sapiens intersecin long isoform (ITSN) mRNA.
 ACCESSION
 AK027846
 VERSION
 AK027846.1 GI:14042823
 KEYWORDS
 oligo capping; fis (full insert sequence).
 SOURCE
 Homo sapiens placenta cDNA to mRNA, clone_1lb:PLACE1
 clone:PLACE1010942.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
 AUTHORS
 1 Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
 Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,
 Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,
 Masuho,Y. and Kanehori,K.
 Masuho,Y. and Kanehori,K.
 NEDO human cDNA sequencing project

TITLE
 JOURNAL
 UNPUBLISHED
 2 (bases 1 to 2131)
 Isegai,T. and Otsuki,T.
 REFERENCE
 AUTHORS
 Direct Submission
 JOURNAL
 Submitted (10-MAY-2001) Takao Isegai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert construction;
 Research Association for Biotechnology; cDNA library construction;
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of Virology, Institute of Medical Science,
 University of Tokyo.

COMMENT
 FEATURES
 source
 1..2131
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE1010942"
 /tissue_type="placenta"
 /tissue="placenta"
 /clone_id="PLACE1"
 /note="cloning vector: pMB18SFL3"

BASE COUNT
 ORIGIN
 691 a 448 c 572 g 420 t

Query Match 34.6%; Score 1259.2; DB 9; Length 2131;
 Best Local Similarity 81.2%; Pred. No. 26-274;
 Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

1 ATGGCTCAGTTTCCACACCTTTGGGTGAGCTGATGTCGGGCGCAATCTGGAG 60
 347 ATGGCTCAGTTTCCACACCTTTGGGTGAGCTGATGTCGGGCGCAATCTGGAG 406
 61 GAAAGGCGCAAGCATGACGAGATTCCTAGCTGAGCGGATAGCGGATTTTACT 120
 407 GAAAGGCGCAAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 466
 121 GGTATCAAGCGGAGAACTTTTTCATCTGGGTGATCTGAGCTGATCTGAGCA 180
 467 GGTATCAAGCGGAGAACTTTTTCATCTGGGTGATCTGAGCTGATCTGAGCA 526
 181 ATATGGCGCTAGCGAGATGATGATGATGATGATGATGATGATGATGATGAT 240
 527 ATATGGCGCTAGCGAGATGATGATGATGATGATGATGATGATGATGATGAT 586
 241 GCGATGAGCTTATCAAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 300
 587 GCTATGAACTTATCAAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 646
 301 GTCATGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 360
 647 GTCATGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 691
 361 GCTAGATGCGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 420
 692 ----- 691
 421 GGAATGCTCCACCTTATGATCTCTGTCCTCCAGCAGAGTGGCTCCCTGGTAC 480
 692 ----- 691
 481 GGGGCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 716 GGGGCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
 541 AAGAGTTCTTCTTCAAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTACGA 600
 776 AAGAGTTCTTCTTCAAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTACGA 835
 601 GCACATTCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 836 GCACATTCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 895
 661 TCAAGGCTGAAATACAGGAGTATTCAACAGCCAGCAAAAATAATGATGAGCACTTA 720
 896 TCAAGGCTGAAATACAGGAGTATTCAACAGCCAGCAAAAATAATGATGAGCACTTA 955
 721 AAGAGTTCTTCTTCAAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTACGA 780
 956 AAGAGTTCTTCTTCAAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTACGA 1015
 781 TCAATATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 840
 1016 TCAATATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1075
 841 CTAGCTATGACCTAATTGATGATGATGATGATGATGATGATGATGATGATGAT 900
 1076 CTAGCTATGACCTAATTGATGATGATGATGATGATGATGATGATGATGATGAT 1135
 901 CCGAATATACCT 960
 1136 CCGAATATACCT 1195
 961 AGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 1196 AGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1255
 1021 C---CAGAGAGAAAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1077

1256 CAATTAGAAAAGAAATTACTGTAACTTTGAAAGATTAAGGAGAACTTTGAAACGT 1315
 1078 GGCAGTGTGAGCTGAGAGACCGCGCAAGCGCTTTGGAGCAGACGCAAAAGCAG 1137
 1316 GGCAGTGTGAGCTGAGAGACCGCGCAAGCGCTTTGGAGCAGACGCAAAAGCAG 1375
 1138 GAGCGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAG 1197
 1376 GAGCGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAG 1435
 1198 CAGAGGCGCAAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 1436 CAGAGGCGCAAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1495
 1496 CAGAGGCGCAAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1555
 1318 CTGAAAAGCGCAGCAACTTTGAAATGGGAACTGAACTGAACTGAACTGAACTGAA 1377
 1556 CTTAAAAGCGCAGCAACTTTGAAATGGGAACTGAACTGAACTGAACTGAACTGAA 1615
 1378 AGAACAAG 1437
 1616 AGAACAAG 1675
 1438 GAGTTAGAGCTCTGAAATGCAAAAAGCAGTACAGTACAGAGAGAGAGAGAGAG 1497
 1676 GAAATAGAGCTCTGAAATGCAAAAAGCAGTACAGTACAGAGAGAGAGAGAGAG 1735
 1498 TGTGAGCTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 1736 TGTGAGCTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
 1558 ATTGCTGAATACCCACTTACAGCAGCAGTTCAGAGATTCAGCAATCTTGAAGA 1617
 1796 ATTGCTGAATACCCACTTACAGCAGCAGTTCAGAGATTCAGCAATCTTGAAGA 1855
 1618 CTTATTCAG 1677
 1856 CTTATTCAG 1915
 1678 CATAGAGCTGCTCTTCTTACCTCAAAAAGCCTTGAAGCAAAAGAGCTGCGCAG 1737
 1916 CACAGAGATTCACCTTCTTACCTCAAAAAGCCTTGAAGCAAAAGAGCTGCGCAG 1975
 1738 CAGCTCCGGAG 1797
 1976 CACCTACAG 2035
 1798 GATGTTTCAACCAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
 2036 GATGTTTCAACCAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
 1858 AAGCAGAGCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893
 2096 AAGCAGAGCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

RESULT 11
 HSB61166 HSB61166 3241 bp mRNA linear PRI 23-JUL-1996
 LOCUS Human SH3 domain-containing protein SH3P17 mRNA, complete cds.
 DEFINITION U61166
 ACCESSION U61166
 VERSION U61166.1 GI:1438932
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 3241)
 AUTHORS Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. and


```

QY 3471 CGATTACCCGCCAGACGATACGAACTAGCTTACGAAAGCCAGATCATCAAGCT 3530
DB 1428 CGACTACACCCGCCAGAAATACGATGAGCTGGCTTCAACAAAGGGCCAGATCATCAAGCT 1487
QY 3531 CCACACCAAGAGAGACCCGGAGCTGTGGAAAGAGAAAGTGTGGGCAAGTGGGCTCTT 3590
DB 1488 CCTCAACAGAGAGAGACCTTCACTGTGGAGAGAGAGATGACACAGTGGGCTCTT 1547
QY 3591 CCATCCAAATTAATGTAAGCTGACCAACAGACATGAGCCCGACCAATGA 3642
DB 1548 CCATCCAAATTAATGTAAGCTGACCAACAGACATGAGCCCGACCAATGA 1599

RESULT 12
AF169621 1133 bp mRNA linear ROD 12-AUG-1999
LOCUS AF169621
DEFINITION Mus musculus Ees1 protein mRNA, partial cds.
ACCESSION AF169621
VERSION AF169621.1 GI:5731280
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Tsymba, L.O., Kvasha, S.M., Skripkina, I.Y., Anoprienko, O.V.,
JOURNAL Slavov, D., Tassone, F., Rynditch, A.V. and Gardiner, K.
AUTHORS Mouse homologues of human chromosome 21 genes
REFERENCE 2 (bases 1 to 1133)
JOURNAL Unpublished
TITLE Direct Submission
AUTHORS Submitted (15-JUL-1999) Department of Molecular Oncogenetics,
JOURNAL Institute of Molecular Biology and Genetics of National Academy of
FEATURES Sciences of Ukraine, 150, Zabolotnogo str., Kiev 252627, Ukraine
source
1. .1133
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="15-day embryo"
<1. .749
/feature="similar to Homo sapiens intersecin"
/codon_start=3
/product="Ees1 protein"
/protein_id="AAB48848.1"
/db_xref="GI:5731281"
/translation="VRKSTIDTPTSPASIKRVASPAKPAIPGEFFIAMTYESS
EGDILTFQGDVIVTKDGMWGTVDKSGVPSNVRLKDSGSGTAGTSLGK
KEPIAQVIVASVATPEQLTAPQQLIIRKKNPGWMEGELQARGKROGWFPANY
VTLSPGSKITPTLPTAVQPAVCQVIMYDYANDDELAFSGQIINVNKEDP
DMWGEVSGQVELPSPNVYKLTMDPQSQ"

BASE COUNT 316 a 276 c 285 g 256 t
ORIGIN
Query Match 20.6%; Score 749; DB 10; Length 1133;
Best Local Similarity 100.0%; Pred. No. 8e-159;
Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2894 CCCTAAGGAATCCACAAGATGATACGCTTACTGAAAGTCTGCTAGTCTTAAAGA 2953
DB 1 CCCTAAGGAATCCACAAGATGATACGCTTACTGAAAGTCTGCTAGTCTTAAAGA 60
QY 2954 GAGTGCCTTCCCGCGCGCCAGCCAGCAATCCCGAGAAAGATTATTCGATGTACA 3013
DB 61 GAGTGCCTTCCCGCGCGCCAGCCAGCAATCCCGAGAAAGATTATTCGATGTACA 120
QY 3014 CATACAGAGATTCTGAGCAAGAGATTAACTTTGACAGAGGGAGTGTATTGGTTA 3073
DB 121 CATACAGAGATTCTGAGCAAGAGATTAACTTTGACAGAGGGAGTGTATTGGTTA 180
QY 3074 CCAGAAAGATGGTGTGAGCGGAAACGGTGGCGACAAAGTCCGAGTCTTCCCTT 3133

```

```

DB 181 CCAGAAAGATGTGACTGTGACGGGAAACGTTGGCGCAAGTCCGAGTCTTCCCTT 240
QY 3134 CTAACCTATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACTGTGGAAAAACAGAGATT 3193
DB 241 CTAACCTATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACTGTGGAAAAACAGAGATT 300
QY 3194 TAGGAAAAAACTGAAATATGCCCCAGGTTATTGCTTCTTACGCTGCTACTGTCCGAC 3253
DB 301 TAGGAAAAAACTGAAATATGCCCCAGGTTATTGCTTCTTACGCTGCTACTGTCCGAC 360
QY 3254 AACTCACCCCTGCTCTGGGAGCTGATTCGATTCGGAAGAAAGAACCCAGGTGATGT 3313
DB 361 AACTCACCCCTGCTCTGGGAGCTGATTCGATTCGGAAGAAAGAACCCAGGTGATGT 420
QY 3314 GGAAGAGAACTGCAAGCTTCGAGGAAAAAGCCGAGATAGGCTGTTCCAGCAATT 3373
DB 421 GGAAGAGAACTGCAAGCTTCGAGGAAAAAGCCGAGATAGGCTGTTCCAGCAATT 480
QY 3374 ATGTCAAACTTCTAAGCCCGGAAACAGCAAAATCACCCCACTGAGCTACCAAGACG 3433
DB 481 ATGTCAAACTTCTAAGCCCGGAAACAGCAAAATCACCCCACTGAGCTACCAAGACG 540
QY 3434 CAGTGCAGCCAGCAGTGTGCGAGGTGATCCGAGTGTACATTAACCGCCAGAACGATG 3493
DB 541 CAGTGCAGCCAGCAGTGTGCGAGGTGATCCGAGTGTACATTAACCGCCAGAACGATG 600
QY 3494 ACGAATTAACCTTACGAAAGCCAGATCATCAAGCTCTTCAACAAAGAGAGACCCGACT 3553
DB 601 ACGAATTAACCTTACGAAAGCCAGATCATCAAGCTCTTCAACAAAGAGAGACCCGACT 660
QY 3554 GGTGAAAGAGAAAGTCAAGTGGGCAAGTGGGCTCTCCATCAATTAATTAAGTGA 3613
DB 661 GGTGAAAGAGAAAGTCAAGTGGGCAAGTGGGCTCTCCATCAATTAATTAAGTGA 720
QY 3614 CCACAGACATGAGCCCGACGACCAATGA 3642
DB 721 CCACAGACATGAGCCCGACGACCAATGA 749

RESULT 13
AF248540 5828 bp mRNA linear PRI 12-APR-2000
LOCUS AF248540
DEFINITION Homo sapiens intersecin 2 (SH3D1B) mRNA, complete cds.
ACCESSION AF248540
VERSION AF248540.1 GI:7542782
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Selfert, M., Engel, M. and Welter, C.
JOURNAL Intersecin 2 (SH3D1B), human homolog of mouse Ees1 protein
REFERENCE 2 (bases 1 to 5828)
AUTHORS Selfert, M., Engel, M. and Welter, C.
JOURNAL Unpublished
TITLE Direct Submission
AUTHORS Submitted (23-MAR-2000) University of Saarland, Institute of Human
JOURNAL Genetics, Building 68, Homburg 66421, Germany
FEATURES
source
1. .5828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="12p"
/cell_line="HEL"
1. .5828
/gene="SH3D1B"
7. .5052
/gene="SH3D1B"
/feature="adaptor protein; similar to Mus musculus ESE2
protein"
/codon_start=1

```

```

/product="intersectin 2"
/protein_id="AAF63600.1"
/db_xref="GI:7542783"
/translation="MASEYQKLRITMAQPTAMNGGNMMAITSEERTKHDPQFNL
KPSGGYITGDARNFPLDGLPAPVLAELMALSLNKGKDDQEFISAKLIKLIQ
GQQLPEVLPIMKQPMFSPILISAFGMSPLNLSIQPLPAPITLSLSATSTNL
PPLNMPPTLVPSVSTSLPNGTASLIQPLIPYSSSTLPHSSSYSLMGSGASIQK
AQLSLDSSSTSTSLNGSPKTSSEMAVPOPTRLKROKFTLMDKMSGYLTS
FOARALLOSLSOTLATITWLDVDDGQLKEEFLIHLNHLMDKAKGQPLPLP
PELVPSFRGKQIDSLNGTLPSYOKMOEEPOKRLPTFEDRDKAVENGMLEER
RQALMEQOQABERKAQKEKEERKQELQEQMKQLELEKLEKQELEREREE
RKQDIEREAKQELERQRLERIRQELQNKQROBEITVLNKSQKXLIYLA
LNKHQOISGRLOJELRKQTELEKQELQELQELQELQELQELQELQELQEL
PEKOLNERIKMOFSNTPDSGSLHKLKLEKELCORLEQDQDALEKEDASLSM
DSFNOLKELRETYNTOALAEOLYKIRKDLKEERLEKLEKLEKLEKLEKLEK
GKENLMEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK
SVLVNRYALYFEARNDHMSFNSDITQVDEKTVGEQGLYSGFQNGFQPNYVE
KMPSENERAVSPKXALPPTVLSLSTSSPSLNOBPATYONVSENLTVNYS
WOKKSAPTRTVSPGSVPDIHQGVENVNKAQALCSMTAKDNHLNFSKHDIIYVLEQ
QENMFGEHGGRGWPKSYKIIPGSEVKEEPEALYAAVNPKNPTSAYSVGEYIA
LYPVSVPGLDTEFEHELVTOKDGMWTSIGDRGIPPSNVVMPKXOESFSA
KSGASNKPEIQAUTSAVYASGSRSLSTAPGOLLILIKKNTSGMGOELARCKKROK
GMPASHVKLGPSSERATPAHPVCOYIANTDYANNEDELSTSKOLLINVMKDP
DMQOGEINGVGLPFSNVVMTTSDPSQWQCADQLDTPQIERKQOYIHELIOT
BERYADQLVVEVFOKMASSGLTEBEMALIFVNMKELIMSTKLIKALRYKKTG
GEMPVOMIGDILAEILSHMOAYIFCSQNLGAALLOQKEDEDTDFEFLKIASDP
RCKMPLSSFLKPMORITRYPLIRSLILENTPSHADHSLKALAELEKQVNE
GVREKENDLEMIQAHVOCGLAEOLFNELTJCTGPRKILHSGKLYKTKSNELG
FLNDFLLITVWKOFAVSSSEKLFSSKSNAGKMYRTPFLNEVLYKLETPDSSDE
PVFHLSHIDRYVTLKTDNIERTAWQKIKASEQYIDTERKEKAVQAQASQTSGL
GLMWVIAETELKAKCPNGKSNPYCELSMGSOSTTITIDITLNPKNFQOPFIQ
LYOVLCITLFDPDQFSPDFLGRTEIIVAKIRTEQBSKGMTRRLILHEVPTGVWV
RFDQLFPQKTL"
BASE COUNT      1867 a      1233 c      1337 g      1391 t
ORIGIN
Query Match      19.8% Score 721; DB 9; Length 5828;
Best Local Similarity 53.9%; Pred. No. 1.9e-152;
Matches 1998; Conservative 0; Mismatches 1430; Indels 279; Gaps 14;

```

```

Db 523 GTTAGCACAATCATCATATCCAAATGGAACCGCAGCTCATTTAGGCTTTACCATTTCTT 582
Oy 508 CTTGCGTTTGGCATCTCTGACGACATAGCGCAAGAGTCTCTTCTTACGACATCTGCT 567
Db 583 TATCTTCTTCAACATTGCTCTCATGGGTATCTTATATGCTGATGATGGAGATTTTGA 642
Oy 568 CCAAGGTACAAATTTAAACATAAGTTACAGAAAGCAATCATTCGATGCGCAGCGCC 627
Db 643 GGTGCTAGTATACAGAAAGGAGCTGCTGATTTAGATCTAGTACCTCAACTTCC 702
Oy 628 CTTCCAG-----CAGCAAAATGGGCTGTGCTT 654
Db 703 TCGACTCTTCACTCTCAGGAACTCACCCAGACTGGGACTCAAGTGGGAGTTCTT 762
Oy 655 CAGTCATCAAGGCTGAATACAGGAGTTATTCACAGCCAGCAAAATATAGATGGA 714
Db 763 CAGCTTCAAGATTAAATATTCGGCAAAATTTAATATCTTTCACAAAGATATGATGGA 822
Oy 715 CACTTAACAGGTCCCAAGGAGAACTATTTCTATGCAATCAAGTTTACCCAGGCTCAG 774
Db 823 TATCTCAGGTTTTCAGAGTAAAGTATGCCCTTCTTCACTCAATCTTTCTCAACTCAG 882
Oy 775 CTGCTTCAATATGGAATCTTTTCTGACATTGATCAAGATGGAATCTCATCGAAGAA 834
Db 883 CTGCTCTATTTGGACTCTGCTGAGAGTGTGATGAGACGCTTAAAGCAGAAAG 942
Oy 835 TTTATCTAGTATGACCTAATGATGTTGCAATGTCTGATGACGACGACGCGCCGTC 894
Db 895 CTGCTCAGAAATACCTCCTCTCTTCTTCAAGAAAGTTGCTCCGCGAGTGGATGCC 954
Oy 1003 TTACTCTCAGGCTTGTTCCTTCATCTTTTCAAGAGAAAGCAAAATGATTCATTAAT 1062
Db 955 GTCAATAGCTTTCTTCTGCTGATCAAGGCTGCTAGAGAGCCGTGTCAAGAGATGAG 1014
Oy 1063 GGAATCTGCTTATATCAAA-----AATGCAAGAA 1095
Oy 1015 CAGAGCCAGAGAAAGAAAGCTCTGTGACATTTGAATGAAGAGCGGAGAACTTGAG 1074
Db 1096 GAGAGCTTCAAGAAATTTACAGTTACTTTTGAAGACAAAGAAAGCCACTATGAG 1155
Oy 1075 CGAGCAGTGTGAGCTGAGAAAGCGCCGCAAGCGCTTTGAGCAGCAGCGCAAGAG 1134
Db 1156 CGAGGAAACATGAGAGCTGAGAAAGCGCAAGCGCTTATGAGAGCAGCAACAAAGGAG 1215
Oy 1135 CAGAGCGGTGCTCAGCTGAGAGCGCGGAGCAGAGAGAGAAAGCGGAGCGCCAG 1194
Db 1216 GCAGAACGTAAGGCCAGAAAGAAAGAGAGAGAGAGAGAGAAAGAGAAATTTCAA 1275
Oy 1195 GAGCAGAGGCGCAAGCGGAGCTGAGCTGGAAGAGAGCTGAGAGAGAGCGGAGCTG 1254
Db 1276 GAAACAAGATGAAGAAACAACTTGAATTGAAAAACGCTTAGAGAGCAACGGGAATTG 1335
Oy 1255 GAGCGCAGCAGAGAGAGAGAGAGAGAGATTCAGAGAGCGCGAGCGCAAAACGG 1314
Db 1336 GAGAGACAAAGAGAGAAAGAAAGAGAAAGACATAGAAAGACAGAGGACGCAAAACAG 1395
Oy 1315 GAACTGAAAGGACGCAAACTTGAATGGGAACGAAACGGAGACAGAACTCTGAAT 1374
Db 1396 GAACTTGAACGACAACTGCTTAGAAATGGGAGAAATTCGGCAGACGAGCTTCTCAAT 1455
Oy 1375 CAGAGAAACAAGAGAGAGAGAGAGAGAGCTGAGAGGCAAGAGAGAGAGAGAGAG 1434
Db 1456 CAAGAGATTAAGAAACAAGAAATTTGTCAAGTTAACTTTAAAAAGAAATCTTCAT 1515
Oy 1435 TTTGATTGAAGCTCTGAATGACAAAGAAAGCATGCTAGAGAGAAATCTTCAAGATATC 1494
Db 1516 CTTGAGTTGAAGCACTGAATGAGCAAAACATCAGCAGATTCAGGACAGACTTCAAGATGC 1575
Oy 1495 AGGTGCACTGGCAACCCAGAGGCAAGAAATTAAGAGCAGAAACAGTCTAGAGACTTA 1554

```

Db 1576 CGACTCAAAAAGCAAACTCAAAAAGCTGAGTGGAAAGTTCTGGATTAAGCAGTGTGACTTG 1635
 QY 1555 AGAATGCTGAATATCACCCTTAACGACGACAGTGGAGGAATCTCAGCAAAATGCTTGA 1614
 Db 1636 GAAATTAATGGAATCAAGCACTTCAACAGAACTTGAGAAATATCAGATTAAGCTTATC 1695
 QY 1615 AGACTTATTCAGAGAAACAGATATCTCAGTGCAGTGTAAACAGTCCAGCAAGACAGT 1674
 Db 1696 TATCTGGTACTTGAGAACATATTAATTAAGAAATTAATAACATGCGAGTTACAGTAAC 1755
 QY 1675 TTGCTATGAACTGCTTCTTACCCTCAAAAGCCCTGGAACCAAGAGAGCTGGCCCG 1734
 Db 1756 ACACTGATTTCAAGGGTCAAGTTTCTCATTAATAAAATCATTAAGAAAGGAATTAATGC 1815
 QY 1735 CAGCAGCTCCGAGAGCAGTGCAGAGTGGAGAGAGACCAAGTCAAAAGCTGCAGAG 1794
 Db 1816 CAAGAATTTAAAGACAGTTAAGATGCTTGAAAAAGAAATCGATCTTAAGCTGCAGAA 1875
 QY 1795 ATTGATTTTTCACAAACCAAGCTGAAGAACTGAGAGATATCATGCAAAACGCAACTC 1854
 Db 1876 ATGGATTTCTTTAAACATCAATAAGAACTGAGAGAAACCTTAACAACACGCGAGTTA 1935
 QY 1855 CAGAGCAGAGGTCCTCGAGGAGCGCGACTGAGAGAGAAAGAGAGAGAGAGAGC 1914
 Db 1936 GCCCTTGAACAGCTTTATTAAGATCAAACTGACAAAGTTGAAGAAATTTGAAGAAAGA 1995
 QY 1915 CTGAGTTAGAGAAAGCAAAAGGAAGACGCTCAGAGACGAGTTGAGAAAGGCAAGCAA 1974
 Db 1996 TTGAACTAATGAGAAAAAGAACTAG-----AAGATG 2029
 QY 1975 TTGCTGAGCATGTGCAAGAGAGAGCAGCAGCGCCCGGAAACCCCAAGAGAGAGC 2034
 Db 2030 AGGCTGCAAGAAAGCAAGCAAGAAAGAAACCTTAATGAAAGAAATCTTAGAAAG 2089
 QY 2035 AGACTGAAGAGGAAAGCAAGTGTCAAGAGAAAGAGCGGAGAGAGAGCCAGCCGAA 2094
 Db 2090 AGGAAGAAAGAAACAAAGACGACTCCAGAGAAAGAAACAAAGAAATTTCAAGAAAG 2149
 QY 2095 ATGCAAGACAGCAGAGTCCGCTTTCCATCCGCATCAGAGAGCAGCTAAGCTGGCCACC 2154
 Db 2150 AGGAACGGAAGCTGAGAGAA-----AACAACTAAGATTAAGATATTTTGAAGCT 2202
 QY 2155 CAGGCACTCTGTCTACACAGAGAAAGCGCCCTTACCATTTCTGACAGAGAGATGTA 2214
 Db 2203 GAGGAGAAAAAGCTGAGACAGCTAGT----- 2230
 QY 2215 AAAGTGTATTTACCGAGCGCTGTACCCCTTTGATCCAGAACTCAAGTGAATCACC 2274
 Db 2231 -TTTTGTGAATTAATAGACATTAACCTTTGAAGCAAGAAACATGATGATGAGT 2289
 QY 2275 ATCCAGCCAGAGATATAGTCATGTGATGAAAGCCAGACTGAGAGCCAGATGGCTT 2334
 Db 2290 TTTAATTTCTGAGATTAATTTCAAGTTGATGAAAAACGCTAGAGAACTGTGTTGGCTT 2349
 QY 2335 GAGAGAGCTGAAAGGAGAGAGCGGATGTTCCCTGCAAACTATGCAAGAAAGATTTCA 2394
 Db 2350 TATGATGATTTTCAAGGAAATTTTGGCTGTTCATGCAATATATGAGAAAAAATGCA 2409
 QY 2395 GAAATAGAGTTCCCACTCCAGCCAAACAGTGCAGATGCAATCTGCCCCCTGCCCCC 2454
 Db 2410 TCAAGTGAATAAT-----GAAAAAGCTGTATCTTCAAGAAAGGCTTACTCTCTCT 2460
 QY 2455 AAATGCTCTGCGTGAAGCCCTGCTCTTTGCAAGTGAATCTTTCTGAGCCCTCACA 2514
 Db 2461 ACAGTTTCTTATCTGTACTCAACTCTCTGAAACCACTTTCTTCAATCAACAGAGA 2520
 QY 2515 ACCCCCAACACTGGGAGACTTCAGTTCCAGTGGCCCAAGCAGCTCAACAGAAAGCCA 2574
 Db 2521 TCAGTACTGATTAATCAAAA----- 2540
 QY 2575 GAAACGACAACTGGGATAGTGGGCGGCTCAGCCTTCTCTGACGTAAGTCTGCTGGC 2634
 Db 2541 -----TGATCTTTTTCAAACCTAAGTGAATATCA 2571

QY 2635 CAGTTAGGCGAGATCAGCTTATACCCAGCCACAGCAGCTGCTCTCCCATCTCC 2694
 Db 2572 TCATGCGCAGAAAAATCAGCTTCACTCGAATGTGTCCCTGG--ATCTGATACACT 2628
 QY 2695 GTCTGGCGCAGGGTGAAAAGGTGAAAGGCTACAGACGCAAGCCCTGTATCTCTGAGA 2754
 Db 2629 ATTCTAGGACAGGAGCAAGGTGTGAAAACTTAATAACAGAGCCCTTGTCTCTGACT 2688
 QY 2755 GCCAAAAAGCAACCACTTAATTTTAACAAAAGTACGTCATCACTGTTTGGAAACAG 2814
 Db 2689 GCAAGAAAGATTAACCACTTCACTTCAAAACATGACATTAATCTGTCTTGGAGAG 2748
 QY 2815 CAAGCATGTGTGTGTAAGAAATCAAGTCAAGAGGTGTGTTCCCAAGCTTAC 2874
 Db 2749 CAAGAAATTTGTGTGTGGAGAGTGCATGAGAGAAAGAGATGTGTTCCCAATCTTAT 2808
 QY 2875 GTGAAACTCATTTCAAGGCGCCGTAGAGAAATCCACAGCATCCATCTGCGCTACTGAA 2934
 Db 2809 GTCAAGATCATTTCTGGAGAGTAAAGTAAACGGGAA-----GAACCAAGACT 2856
 QY 2935 AGTCTGCTAGCTTAAGAGAGTGTGCTCCCGGCGCCCAAGCCAGCTATCCCGAGAA 2994
 Db 2857 TTGTATGACAGCTGTAATTAAGAAACCTACCTCGCAGCTTATTCAG-----TTGGAGAA 2910
 QY 2995 GAGTTTATTCGATGTACATACAGAGTTCTGAGCAAGAGATTTAATCTTTACGAA 3054
 Db 2911 GAAATATTTGACCTTTTCCATATTTCAAGTGTGAACCTGGAATTTGACTTTTACAGAA 2970
 QY 3055 GGGAGTGTGATTTGTGTTAACCAAGAAATGTGATCTGTGAGACGGAGCGGTGGCGAC 3114
 Db 2971 GGTGAAGAAATATTTGTGTGACCCAGAAAGATGAGAGTGTGTGACAGAAATATTGGAGAT 3030
 QY 3115 AAGTCCGAGTCTTCCCTTCTAATGTGAGCTTAAAGATTCAGAGCTCTGCACT 3174
 Db 3031 AGAATGGAATTTTTCATCACTATGTCAACCAAGAGATCAAGAGATTTTGGAGAT 3090
 QY 3175 GGTGGAAAAACAGGAGATTTTGAAGAAAAACCTGAATTTCCCGAGTTATTTGCTCTAC 3234
 Db 3091 GCTAGCAAGTGTGAGCATCAATTAATAAACCTGAGATTTGCTCAGGTAACTTACGCTAT 3150
 QY 3235 GCTGTACTGTGTCCGAAACAATCACCCTGTGCTTGGGAGCAGTATTCGATCCGAAA 3294
 Db 3151 GTTGCTTCTGTGCTGAACAACTTAGCTTGACCAAGACAGTTAATTAATTTCTAAAG 3210
 QY 3295 AAGAAACCAAGTGTGATGTGGAGAGAGAACTGAACTGAGAGAAAGCCCGAGATA 3354
 Db 3211 AAAAAATCAAGTGTGTGTGCAAGAGAGTTACAGGCCAGAGAAAAAGCAAGAA 3270
 QY 3355 GGTGTGTTTCCAGCAAAATTAATGTCAAACTTAAAGCCCGGAAACAAAGCAAAATCACC 3414
 Db 3271 GGATGTGTTCTCGCAAGTCAATGTTAACTTTTGGGTCCAAAGTGTGAAGAGCACACT 3330
 QY 3415 ACTGAGCTAACCAAGACCGGACGTCAGCAGCAAGTGTGCAAGTGTGAGTGTACAT 3474
 Db 3331 GCTTTATATC-----CTGTATGTAGAGGATTTGTTATGTATGAC 3369
 QY 3475 TACACCGCCAGAGAGATGAGAACTAGCTTCAAGAAAGCCAGATCATCAAGTCTCTC 3534
 Db 3370 TATGAGCAAAATTAAGAGATGAGCTCAGTTTCTCCAAAGGAGCAACTTAATGTATG 3429
 QY 3595 TCCAAATTAATGAAGCTGACACAGACATGAGACCCCGCAGCAAGT 3641
 Db 3490 TCAAACTACGTTAAGATGACGACAGCTCAGATCAAGTCAAGTGT 3536

RESULT 14
 AB033082
 LOCUS AB033082 5938 bp mRNA linear PRI 11-NOV-1999

DEFINITION Homo sapiens mRNA for KIAA1256 protein, partial cds.
 AB033082
 VERSION AB033082.1 GI:6330910
 KEYWORDS
 SOURCE Homo sapiens brain cDNA to mRNA, clone_11b; pBluescriptII SK plus
 clone:h15293.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (sites)
 Nagase, T., Ishikawa, K., Kikuno, R., Hirose, M., Nomura, N. and
 Ohara, O.
 Prediction of the coding sequences of unidentified human genes. XV.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro
 DNA Res. 6 (5), 337-345 (1999)
 2 (bases 1 to 5938)
 Ohara, O., Nagase, T. and Kikuno, R.
 Direct Submission
 Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology; 1532-3 Yama, Kisarazu, Chiba
 292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,
 URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,
 Fax: +81-438-52-3914)
 FEATURES
 source
 1.5938
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="h15293"
 /cissue_type="brain"
 /clone_lib="pBluescriptII SK plus"
 1.5938
 /gene="KIAA1256"
 /gene="KIAA1256"
 /note="Start codon is not identified."
 /codon_start=1
 /product="KIAA1256 protein"
 /protein_id="BA06570.1"
 /db_xref="GI:6330911"
 /translation="OGLRTTMAAOPFTANNGSPNNMALTSEERTKHDRQFNDLKPSSG
 YITGDARNEFLQSGPAPVLAIEIMALSDLNKDGKDOEFIMAKLILKLGQGLP
 VILPTMKOPPMSPSPISARFGMGSPNLSIPOLPPAPRPISSATSGTNPPLMM
 PPIPLVSVSTSSIPNGTASLIQPLPIPISSSTLPHSSSTSLMGFGGASIOKASLI
 DGGSSSTSTASLNGSPKRTSEMAVPOPTLRKQKQFNTLDSMSGYLSGFGARN
 ALIOSNLSTQTLATITWTLADVDGQLKAEFFLMAHLIDMAKAQPLPLTLPPLVP
 PSFSGKQIDSLINGTILPSYQKQEEPOKPLPTVEDEKAYERGNMELERQALM
 EOOQREARKAOKKEKEMERKOELOEOKKOLLEKLEKRELEROREBERERKDI
 ERREAKOLERORLEMERIRQELINKRDOEIVLNSKKXKLILEALNGKH
 OQISGRLODRUKQOTKTELEVLDKOCLEIMETKQDOELQOETONKLIIVPERKOL
 LNERIKNMFSTNPDGVSLLHKKSLKEKELCQRLKQDLDALEKTAASKLSDNFNN
 OLKELEKRENTQOLAEOLYKIKRDLKEIKERKLELMOKKLEBAARAKAQGENL
 MKELEKEEKKOQLKEKTOEKIQEERKAEKORKDOKTLKAEKRETAASVLV
 YRALYFEARNDHMSFNGSDIIOVEKTVGEGMLYSGFQNGFNGFPCNVEKPS
 ENKKAASPKKALLPTVSLSATSTSEPLSNQPSAIVTQVNSNLTNTVMSKKS
 ATRTVSPGVSPIHGOQVENVKALCALCSWTAKDNHLSKDIITVLEOENMM
 FGEVHGRTGEMFPKSVKIIIPGSEVKEEPEALAAVANKPTSAASVGEVIALPYIS
 SVPEGLDTFEGEELIVTQKDEMTGSIGDSGILPSSVYVRKQESVSGASKGS
 NKKEIAQVTSAAVAGSEQLSLAPQOLILIKNTSGMWGELQARGKRGKGFAPA
 SHVKILGPSERATPAFHPVCOVIALYDVAANDELSFSGQILVNMKDDPDMWG
 EINGVGLPSPNVKMTTSDSQOCADLOTLDMOPIERRQGVHILIOETERRM
 ADIQVVEVFOKMAESGFLTBGMALIVYNNKELINSTKLIKALIRAKKTGGERMP
 VOMIGDILAEELSHQAVIRFCSCQNLGALLOQKDEDTDEPEFLKASPRCKGM
 PLSSFLKEMQIRITRYPILRSILENTPESHADSHSLKALRAEELCQVNEGKEX
 ENDRLEMIQAHQCEGLAEOLIFNSLTICLGRKLHSLGKLYKTSNKEHLGFLFND
 FILITVWVOFAVSGSEKLFSSKNAOAFMYTPIELNEVILKPTDSSSEPFH
 SHIDRYVTLPTDNIINERTAMVOKIKASAOFYIDTEKROEKIYVQASOKTSGIGLMI
 HUIEATBLKACKRGKSNPYCEISKSOSTTTTIDTDLNPKNNRCQFIDILYQV
 LCILFDROQSPDDFLGRTETPAKIRTEOESKGMTRLLHEVTEGSEVWRPDLQ
 LFEOKTLL"

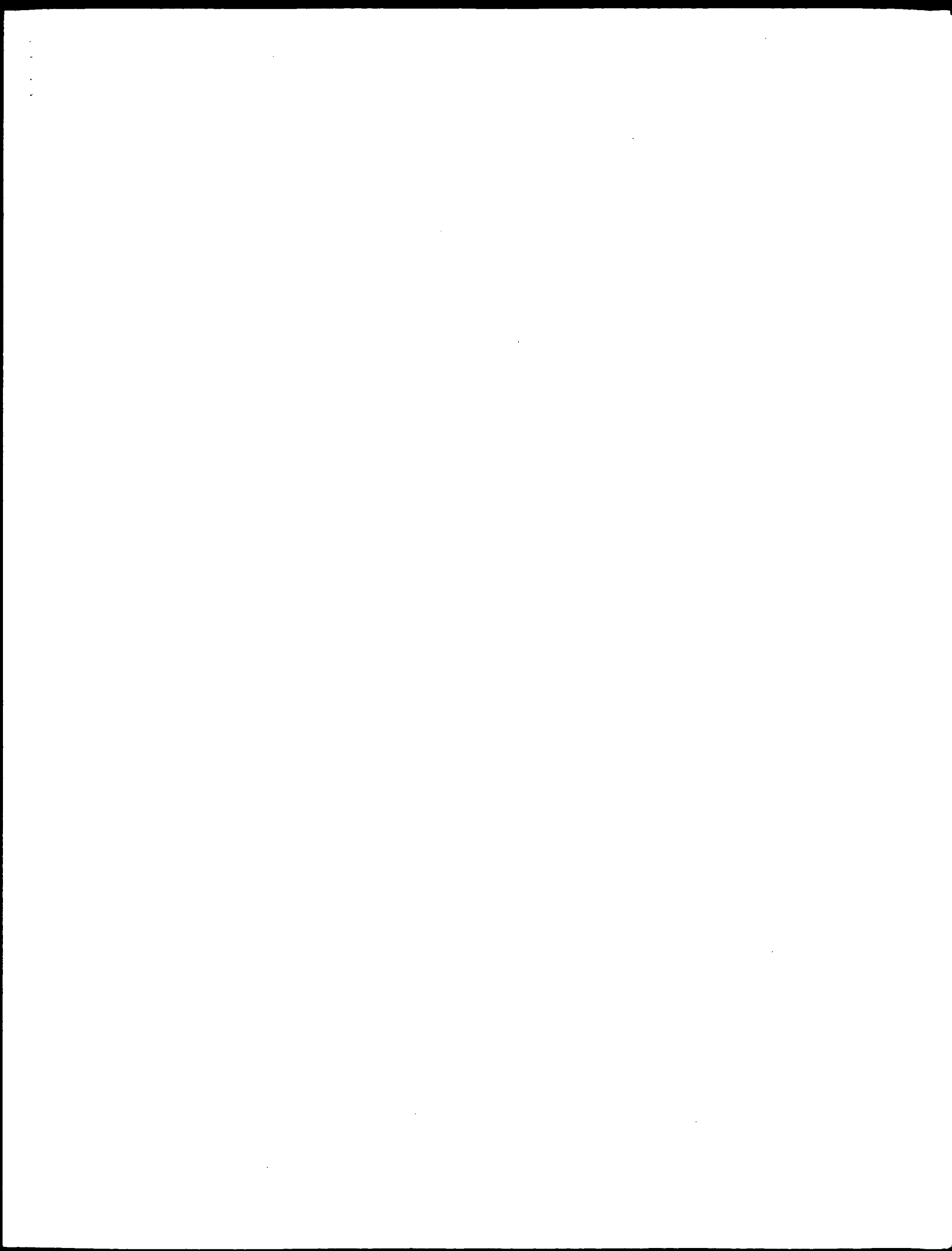
Query Match 19.8%; Score 721; DB 9; Length 5938;
 Best Local Similarity 53.9%; Pred. No. 2e-152;
 Matches 1998; Conservative 0; Mismatches 1430; Indels 279; Gaps 14;
 1 ATGGCTAGTTTCCCAACCTTTGGTGTAGACCTGATGTCTGGCCATTACTGGAG 60
 169 ATGGCTAGTTTCCCAACCTTTGGTGTAGACCTGATGTCTGGCCATTACTGGAG 228
 61 GAAGGGCAGCATGACCCAGCTTCTAGCCTGAAGCCATAGGGGATTTACT 120
 229 GAAGGGCAGCATGACCCAGCTTCTAGCCTGAAGCCATAGGGGATTTACT 288
 121 GATGATCAAGGAGAACTTTTTCCTCAATCTGGTGTAGACCTGATGTCTGGAG 180
 289 GATGATCAAGGAGAACTTTTTCCTCAATCTGGTGTAGACCTGATGTCTGGAG 348
 181 ATATGGGGCTAGGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 349 ATATGGGGCTAGGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 408
 241 GCGATGAGCTTATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 409 GGTATGAACTCATTAACCTGATGATGATGATGATGATGATGATGATGATGATGAT 468
 301 GTCATGAAACGCAACCACTG---GCTATTTCCAGTGCACGACGATTTGTATAGAGG 357
 469 ATATGAAACGCAACCACTG---GCTATTTCCAGTGCACGACGATTTGTATAGAGG 528
 358 ATATGAAACGCAACCACTG---GCTATTTCCAGTGCACGACGATTTGTATAGAGG 390
 529 ATATGAAACGCAACCACTG---GCTATTTCCAGTGCACGACGATTTGTATAGAGG 588
 391 CCTGTCGAATGGGCTCA---TTCCAGTGTGGAATGTCACACCTTATGATCTTCT 447
 589 TCTGGAATTCAGGAGCAACCTTCTCTCTTAATGATGATGATGATGATGATGATGAT 648
 448 GTCCTTCAGGAGCAACCTTCTCTCTTAATGATGATGATGATGATGATGATGATGAT 507
 649 GTTACACATCATCTTACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 708
 508 CCTGTCGAATGGGCTCA---TTCCAGTGTGGAATGTCACACCTTATGATCTTCT 567
 709 TATTCCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
 568 CCAGGTCACAAATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 627
 769 GGTGCTAGTATACAGAAAGCGCAGCTCTGATGATGATGATGATGATGATGATGATGAT 828
 628 CCTCAG-----CAGCAGAAATGGGCTGTGCT 654
 829 TCGACTCTTCACTCTCAGGAACTCACCAGATGAGGAGCTCAGAGTGGGAGTTCT 888
 655 CAGTCATCAAGGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 714
 889 CAGCTCAAGGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 948
 715 CACTTAACAGGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 774
 949 TATCTCTCAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 1008
 775 CTTGCTCAAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 834
 1009 TTTATCTCAAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 1068
 835 TTTATCTCAAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 894
 1069 TTTATCTCAAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 1128
 895 CTGCTTCAAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 954
 1129 TTTATCTCAAGGCTTCAAGCTGAAATAGAGGAGTATTAACAGCAGCAAGCAAAATATAGATGGA 1188

QY 955 GTCATTAAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGCTGACAGATGAG 1014
 Db 1189 GGAAGCTGCTCTCATATCAGAA-----AATGCAAGAA 1221
 QY 1015 CAGAGCCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAAAGCGGAGAACTTTCGAG 1074
 Db 1222 GAGAGGCTCAGAAAGAAATTAACGATTACTTTTGAAGACAAAGAAAGCCAACTTATGAG 1281
 QY 1075 CAGAGCAGTGTGAGCTGAGAAAGCCGCCAAGCGCTCTTGGAGCAGCAGCCAAAGAG 1134
 Db 1282 CGAGGGAACATGAGCTGAGAAAGCCAGCCAGCTTGTATGAGCAGCAACAAAGGAG 1341
 QY 1135 CAGAGCGGTTGGCTCAGCTGTGAGCCGCCAGCAGAGAGAAAGAACGCCGAGCCCGCAG 1194
 Db 1342 GCAAGACGTAAAGCCAGAAAGAAAGAAAGTGGAGCGAAACAGAGAGATTTACAA 1401
 QY 1195 GAGCAGAGGCCCAAGCGCGAGCTGAGCTGAGAAAGCAGCTGAGAAAGCAGCGAGCTG 1254
 Db 1402 GAAACAAGATGSAAGAAACAACTTGAATTAGAAAAACCTTAGAAGACAAAGGAAATTG 1461
 QY 1255 GAGCGGAGCGAGAGAGAGAGAGAGAGAGATGAGAGGCGCGAGCCGCAAAACGG 1314
 Db 1462 GAGAGACAAAGAGAGAGAGAGAGAGAGATGAGAAAGACATGAAAGAGAGAGCAGCAAAACAG 1521
 QY 1315 GAACTGGAAGGAGCAGCAACTTGAATGAGAAACGGAACCGAGAGAGAGAACTCCTGAAT 1374
 Db 1522 GAACTTGAAGAGACAGACGCTCCGTTAGATGAGAGAAATTTGGGAGACAGAGAGCTTCAAT 1581
 QY 1375 CAGAGAAACAAG 1434
 Db 1582 CAAAGAAATGAGAAACAAGAAAGAAATGTGAGTTAACTTAAAGAAAGAAATCTTCAAT 1641
 QY 1435 TTGAGTTAGAGCTCGAATGACAAAGACATCAGCTAGAGAGAAACCTTCAGATATC 1494
 Db 1642 CTGAGTTGAG 1701
 QY 1495 AGGTGTGAGCTGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554
 Db 1702 CGACTCAAAAGAGAACTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
 QY 1555 AGAATGCTGAAATCAACCCCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
 Db 1762 GAAATTAATGAGAAATCAAGCACTTCAACAGAACTTCAAGAAATATGAGATAGCTTATC 1821
 QY 1615 AGACTTATTCAGAGAAACAGATACAGTACAGTAAACAGTAAACAGTAAACAGTAAAC 1674
 Db 1822 TATCTGATCCTGAG 1881
 QY 1675 TTGCAATGAGACTGCTCTTCTTACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAG 1734
 Db 1882 ACACCTGATTCAG 1941
 QY 1735 CAGCAGCTCCGAG 1794
 Db 1942 CAAAGACTTAAG 2001
 QY 1795 ATTGATGTTTCAACCAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
 Db 2002 ATGAGATTCCTTTAACAATCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2061
 QY 1855 CAG 1914
 Db 2062 GCCCTTGAACAGCTTTATTAAGATCAACAGTGAACAGTGAAGAGAGAGAGAGAGAG 2121
 QY 1915 CTGAGGTTAG 1974
 Db 2122 TTGAGATTAATGAG 2155
 QY 1975 TGGCTGAGAGATGAG 2034
 Db 2156 AGGCTGCAAG 2215
 QY 2035 AGACTGAAG 2094
 Db 2216 AGGAAG 2275
 QY 2095 ATGCAAG 2154
 Db 2276 AGGAAG 2328
 QY 2155 CAGGACCTGCTGTACCAAG 2214
 Db 2329 GAGGAG 2356
 QY 2215 AAAGGTATTTATCCAG 2274
 Db 2357 -TTTGTGAATTTAAGAGATTAATCAGCTTGTACCCCTTTGAGAGAGAGAGAGAGAG 2415
 QY 2275 ATCCAGCAGAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2334
 Db 2416 TTTAATTCAGAGATATATTAATCAGCTTGTACCCCTTTGAGAGAGAGAGAGAGAGAG 2475
 QY 2335 GAGGAGAGCTGAAAG 2394
 Db 2476 TATGTAGTTTCAAGAGAAATTTGGCTGTTCATGATGATGATGATGATGATGATGAT 2535
 QY 2395 GAAATGAGGTTCCCACTCCAGCCCAACAGTGCAGATCTGACATCTGCCCTGCCGCC 2454
 Db 2536 TCAAGTGAAT-----GAAAGAGTATCTCCAAAGAGAGAGAGAGAGAGAGAGAGAG 2586
 QY 2455 AAATGCTGCTGCTGAG 2514
 Db 2587 ACAGTTCTTATCTGATCCTCAACTCTGAACTTCTGAACTTCTTCAATCAACAGCA 2646
 QY 2515 ACCCCCAACAGTGGAG 2574
 Db 2647 TCAAGTGAATTTCAAA-----TGATCTTTTCAAACTTAAGTAAATACA 2666
 QY 2575 GAAACGAGCACTGGAGATAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2634
 Db 2667 -----TGATCTTTTCAAACTTAAGTAAATACA 2697
 QY 2615 CAGTTACGAGAGATCAGAGCTTACCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2694
 Db 2698 TCATGGCAG 2754
 QY 2695 GTCTGGCCAG 2754
 Db 2755 ATTATGAG 2814
 QY 2755 GCGAAAAAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAATCAGCTTCTGAAAG 2814
 Db 2815 GCAAG 2874
 QY 2815 CAAGCATGAG 2874
 Db 2875 CAG 2934
 QY 2875 GTGAAGCTCATTTGAG 2934
 Db 2935 GTCAAGATCATCTCGGAG 2982
 QY 2935 AGTCTGCTAGTCTAAAG 2994
 Db 2983 TTGATGAGAGCTGAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3036
 QY 2995 GAGTTTATTTGAG 3054
 Db 3037 GAATATATTTGAG 3096
 QY 3055 GGGAGTGTGATGTTGTTTCAAG 3114
 Db 3097 GGTGAAG 3156
 QY 3115 AAGTCCGAGAGCTTCTTCACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3174

Db 3157 AGAAGTGGATTTTTCATCAAACTATGTCAAAACCAAGATCAAGAGCTTTGGAGT 3216
 Qy 3175 GCTGGAAAACAGGAGATTAGGAAAAAAACCTGAATTTCCAGATTATGCTTCTAC 3234
 Db 3217 GGTAGAACTGTGGAGCATCAATAAAAACTGAGATTGCTCAGTAATCTCAGCATAT 3276
 Qy 3235 GCTGCTACTGTGTCGGAACAACCTACCTGCTCCTGGGAGCTGATTTGATCCGAAA 3294
 Db 3277 GTTGGCTTCTGTTCTGMACTAGCTTTCGACAGGACAGTAAATTAATCTAAAG 3336
 Qy 3295 AAGAACCAGGTGGATGTGGGAGAGAACTGCAAGCTCGAGGAAAAAGCGCAGATA 3354
 Db 3337 AAAAATACAGTGGGTGGGAGAGAGATTACAGCCAGGAAAAAGGACAGAAAA 3396
 Qy 3355 GGGTGGTTTCCAGCAAAATTATGTCAAACTTCTTAAGCCCGGAAACAGCAAAATCACCCCA 3414
 Db 3397 GGATGGTTTCCGCGCAGTATGTTAACTTTGGGTCGAAGTAAAGAAAGCCACACT 3456
 Qy 3415 ACTGAGCTACCAAGACCGCAGTGCAGCAGAGTGTGCTGATGGGATGTACAT 3474
 Db 3457 GCTTTTCATC-----CTGATGTGAGTATGCTATGTATGAC 3495
 Qy 3475 TACACCCCGCCAGACGATACGATAGCTTCCAGCAAAAGCCAGATCATCACTCTC 3534
 Db 3496 TATGCAAGCAAAATTAATGAATGATGAGCTTCTCCAAAGGACAACTCATTAATGTATG 3555
 Qy 3535 AACCAAGAGACCCCGAGCTGTGGAAAGAGAGATCAGTGGCAAGTTGGGCTTTCCCA 3594
 Db 3556 AACCAAGATGATCTGATTTGGTGGCAAGAGATCAACGGGATGATGCTTTCTTCT 3615
 Qy 3595 TCCATTATGTAAAGCTGACCAAGACATGACCCAGCCAGCAATG 3641
 Db 3616 TCAACTAGCTTAAGATGACGACAGACTCAGATCCAGTCAACAGTACAGT 3662
 RESULT 15
 AF182199 4557 bp mRNA linear PRI 17-AUG-2000
 LOCUS Homo sapiens intersecin 2 short isoform (ITSN2) mRNA, complete
 DEFINITION cds.
 ACCESSION AF182199
 VERSION AF182199.1 GI:7329077
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 4557)
 Pucharcos,C., Escivill,X. and de la Luna,S.
 TITLE Intersectin 2, a new multimodular protein involved in
 clathrin-mediated endocytosis
 JOURNAL FEBS Lett. 478 (1-2), 43-51 (2000)
 MEDLINE 20382722
 PUBMED 10922467
 Pucharcos,C., Escivill,X. and de la Luna,S.
 REFERENCE 2 (bases 1 to 4557)
 TITLE Direct Submission
 AUTHORS Submitted (01-SEP-1999) Medical and Molecular Genetics Center,
 Journal Cancer Research Institute, Avia, Cascelldeldefels Km 2.7,
 L'Hospitalet de Llob, Barcelona 08907, Spain
 FEATURES
 source
 1..4557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p23"
 1..4557
 /gene="ITSN2"
 241..3987
 /gene="ITSN2"
 /note="EH-SH3 multidomain protein"
 /codon_start=1
 /product="intersectin 2 short isoform"

/protein_id="AAF59904.1"
 /db_xref="GI:7329078"
 /translation="MAQFPPTAMNGGPNMMAITSEBRTXDRPDNI.KPSGGYITDQA
 RNFLOSGLPAPLAEIWMALSDIKDQKMDQFSLAMLI.LKIKQGOOL.PVLPLPIM
 KQPMFSPPLISAPFGSMNPINLIPQPPAPRIPISATSGTLPLPMLPPLVPS
 VTNSSLPNPTASLIQPLPIPYSSSTLPHSSSILMMGCGGAS.IOKAOSLIDIGSSSS
 TSSTASLSGNSPRTGTSEMAVFPQTRFLKRLQRLNLDLDSMSGYLSGFQRRNALDSNL
 SQTQATITWTLADVDGDLKAEFLILAMHLIDMAKAGQPLPLTLPLPELVPSPFRGK
 QIDSINGTLPSYOKMOEBEPQKLLPVTEDEKRAYENRNMLEKROQLMCOQOREA
 ERAKOEKEMERKOELOEOMKOLELEKORLEBERKOEEREDIERBAAL
 OELERORLEMERIRROELNKNRBOELIYVNLKKKHLLELALNKHOOISGRIL
 QVRLKQKQKTELEVLDDKCDLIMELEIKOLOEIOEYONKLIYVPEKQILNERTKN
 MQPSNTPDGVSBLHKSLLEKELCORLRLQDLALEKELASFLSEDSNNOLCKGM
 DSVLQCLLSLCLNNLFLILKELREYNTQOLLEOLYKILRKULKEIRKLELM
 OKKLEDEARAKAKOKEMLKEMENLEKEEBEOKLEOEKIOEKIOEBERKAEKQK
 DKOTLKAEEKRETSVLVNYRALYFELARNHDEMSFNSGDIIOYDEKTVGEMLYG
 SPQNGKGFPCNVYEMKPSSENEKVASPKKALLPVLSLSATSSBPLSSQIPASVT
 DYONVSFSLVNTVNTSOKKSAFTRVSPSVPIHGQGOYVNLKQALCSTAKDN
 HLNFSKHDIITVLEQDENMWFGEVHGKMFPSKIVKLI.PGSEVREBEELIYAVLNK
 KPTSAVSVGEBYIALPYSSVPEDDLFTTEBEEILVYQKDEMTGSI.GDSGLFPS
 NYVKPQDES.FGSASKSGASNNKPEIAQVTSAYVASGFSOLSLAFGOLILILKNTGS
 WMGSELQARGKKRKQKGFPAHSVXKLIGPSESEATPAFHVCOVILAMYDVAANNBELS
 FSKGOLINMNDPDDPMWQELINGVTGL.PRSYVVMPTTDDSPSQWQCADLOTLDIMOP
 IERKRGYIHEILQTBERTMADQLVVEVWRLLIASRSKICLS"

BASE COUNT 1463 a 919 c 1081 g 1094 t
 ORIGIN
 Query Match 19.6% Score 713.8; DB 9; Length 4557;
 Best Local Similarity 53.4%; Pred. No. 8.2e-151;
 Matches 1989; Conservative 0; Mismatches 1502; Indels 234; Gaps 14;
 Qy 1 ATGGCTCAGTTTCCACACACCTTTGCGGTAGCCTGGATGCTGGGCCATTAACGTGGAG 60
 Db 241 ATGGCTCAGTTTCCACACAGCTATGAATGAGAGGCCAAACATGTGGCTATTAACCTCGAA 300
 Qy 61 GAAAGGCCAAGCATGACCGACAGTTCTTAGCCTGAAGCCATAGCGGATTTATTACT 120
 Db 301 GAAAGCTTAAGCATGACGAGAGATTGATTAACCTCAAGAGGTTTACATPAACA 360
 Qy 121 GGTATCAGCGAGGAACCTTTTTCCTCAATCTGGGTTACTTCAGCCTGCTTAGACAA 180
 Db 361 GGTATCAGCGAGCTTAATTTTTCCTCAATCAAGTCTGCCGCCCTCTGTTTAACTGAA 420
 Qy 181 ATATGGCGCTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTTCATA 240
 Db 421 ATATGGCGCTTATGAGCCTTAACCAAGAGATGGAAGATGATGACGAAGAGTTCTCCATA 480
 Qy 241 GCAATGAAGCTTATCAAACTGAAGCTCAAGATATCAGTCTCCCTTCACACTTCCCT 300
 Db 481 GCTATGAAGCTTATCAAACTGAAGCTTCAAGGCCAACAGTTGCTGTGTTCTCCCTCCT 540
 Qy 301 GTCATGAAGACGACCAACGAGT---GCTATTTCAGTGCACACACATTTGGTATGAGAG 357
 Db 541 ATTATGAAGACACCCCTATGTTTTCCTCAATTAATTTCTGCTGTTTGAATGGAAGAGC 600
 Qy 358 ATTGCTAGCATG-----CCACACTCAACAGCTGTGTCT 390
 Db 601 ATGCCCCATCTGTCCATCTCTCAGACCATGCTCCACAGCTGACCTTAATCATATGTGCT 660
 Qy 391 CTTGTGCCAATGGGCTCA---TTCCAGTTGTTGAATGTCTCCACCTTATGATCTTCT 447
 Db 661 TCTGCACTTCAGGACCAACCTTCTCCCTTAATGATGCCACCTCCCTAGGCTTCT 720
 Qy 448 GTCCCTCAGCAGAGAGTCTCCCTGCTCAAGGGGCTCTCCCTCATACAGCCTCTG 507
 Db 721 GTTAGCACATCATATTATTAACCAATGAACCGCAGTCTATTAGGCTTATTCATTTCT 780
 Qy 508 CTTGCTTTGGCGATCTCTGAGCCACATGCGCAAGAGTCTTCTTTCAGACAGATGTGT 567
 Db 781 TATCTTCTTCAACATGCTCATGGGTATCTTATAGTGTANTAGTGGAGAGATTTTGA 840
 Qy 568 CAGGGTCACAATTAACTAAGTTACAGAGGACCAATCATTTGATGTGCGCAGCGCC 627



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:28:45 ; Search time 650.685 Seconds
(without alignments)
12604.835 Million cell updates/sec

Title: US-09-674-237A-2
Perfect score: 3642
Sequence: 1 atggctcagttccacaccc.....tggaccaccagccagcatga 3642

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq.101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3642	100.0	3723	21	AAZ39009
2	3639.4	99.9	5144	21	AAZ39025
3	3639.4	99.9	5738	21	AAZ39024
4	3630	99.7	5082	21	AAZ39008
5	2856	78.4	5195	20	AAZ34572
6	2831	77.7	5458	20	AAZ34571
7	2491.6	68.4	7435	23	AAZ34573
8	2429.6	66.7	5199	20	AAZ34570
9	2043.8	56.1	3319	23	ABK43498

10	2043.8	56.1	3466	22	AA163825	Human polynucleoti
11	1495	41.0	2079	20	AAZ24573	Human SH3D1A CDNA
12	1483.8	40.7	2131	22	AAK94139	Human Full-length
13	1291.4	35.5	2874	23	AAZ34572	DNA encoding novel
14	1259.2	34.6	2131	22	AAH16578	Human CDNA sequenc
15	1252.8	34.4	3231	20	AAZ34574	Human SH3D1A CDNA
16	1104.2	30.3	1676	22	AAK94611	Human SH3D1A, SRQ ID
17	858	23.6	1389	17	AAI97975	Human SH3P17 gene.
18	721	19.8	5828	24	AAI47247	Allergic disease e
19	713.8	19.6	6103	22	AAK52332	Human polynucleoti
20	677	18.6	4625	21	AAZ39010	Mouse Ees2l full le
21	677	18.6	4975	21	AAZ39027	Mouse Ees2l coding
22	677	18.6	6014	21	AAZ39026	Mouse Ees2l CDNA s
23	666.6	18.3	3593	21	AAZ39011	Mouse Ees2l coding
24	648	17.8	4447	22	AAZ39025	DNA encoding molec
25	485.4	13.3	877	22	AAK91610	Human CDNA 5'-end
26	485.4	13.3	877	22	AAK93179	Human CDNA clone r
27	477.4	13.1	2017	21	AAZ69762	Human ovarian carc
28	477.4	13.1	2017	24	AAH72656	Human carcinoma
29	387.4	10.6	1329	22	AAH15280	Human CDNA sequenc
30	356	9.8	548	22	AA180000	Human polynucleoti
31	356	9.8	568	22	AA163919	Human polynucleoti
32	356	9.8	568	22	AAZ31621	DNA encoding nove
33	356	9.8	568	23	ABK43814	DNA encoding novel
34	340.8	9.4	2067	24	ABQ55007	Human ovarian anti
35	332.4	9.1	2873	17	AAI97979	Human clone 65 gen
36	331.4	9.1	3746	23	ABK43586	DNA encoding novel
37	331.4	9.1	4210	22	AAZ27090	CDNA encoding nove
38	281	7.7	831	22	AAH03435	Human CDNA clone (
39	260	7.1	676	22	AAH08146	Human SH3P18 gene.
40	258	7.1	747	17	AAI97976	Human secreted exp
41	250	6.9	346	21	AAZ44038	DNA encoding novel
42	210.8	5.8	395	23	AAZ84759	Human breast cell
43	198.8	5.5	270	22	ABA46803	Human foetal liver
44	198.8	5.5	270	22	ABA46803	Probe #10273 for g
45	198.8	5.5	270	22	ABA11807	

ALIGNMENTS

RESULT 1
ID AAZ39009 standard; CDNA; 3723 BP.
XX AAZ39009;
AC
XX
XX 28-FEB-2000 (first entry)
DT
XX
XX
DE Mouse Eesl coding sequence.
XX
XX
XX Mouse; murine; Eesl; Ees2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
XX Mus sp.
OS
XX
XX W09955728-A2.
PN
XX
XX 04-NOV-1999.
PD
XX
XX 27-APR-1999; 99WO-CA00375.
PF
XX
XX 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
PA
XX
XX Egan SE, Wang W, Sengar A;
PI WPI, 2000-052802/04.
XX
XX P-PsDB; AAY57444.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 XX
 XX
 PS Claim 6; Page 40-42; 99pp; English.
 CC The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its agonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp15 complex, then binding dynamn to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX
 XX Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;
 Query Match 100.0%; Score 3642; DB 21; Length 3723;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 3642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 541 AAGAGTTCTTCTTCAGCAGATCTGCTCCAGGCTCAACAATTAAACCTAAGTTACAGAG 600
 Qy 601 GCACATCATTCGATGTCGCCAGGCCCTCCAGCAGACAGATAGGCTGTGCTCACTCA 660
 Db 601 GCACATCATTCGATGTCGCCAGGCCCTCCAGCAGACAGATAGGCTGTGCTCACTCA 660
 Qy 661 TCAGAGCTGAAATACAGCAGTTATTCACAGCCACGACGACAAATATAGATGACACTTA 720
 Db 661 TCAGAGCTGAAATACAGCAGTTATTCACAGCCACGACGACAAATATAGATGACACTTA 720
 Qy 721 ACAGTCCCGCAGAGAACATCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780
 Db 721 ACAGTCCCGCAGAGAACATCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780
 Qy 781 TCAATATGAAATCTTTCTGACATTGATCAAGTGAATACTCACTGCAAGAAATTTATC 840
 Db 781 TCAATATGAAATCTTTCTGACATTGATCAAGTGAATACTCACTGCAAGAAATTTATC 840
 Qy 841 CTAGCTATGACCTTAATTGATGTCATGTCAGGACGACCTGCGCCGCTGCTGCT 900
 Db 841 CTAGCTATGACCTTAATTGATGTCATGTCAGGACGACCTGCGCCGCTGCTGCT 900
 Qy 901 CCAGATATCATCTCTCTCTCTTCCAGAAAGATTGCTCCGCACTGGAGTGTCCGTATA 960
 Db 901 CCAGATATCATCTCTCTCTCTTCCAGAAAGATTGCTCCGCACTGGAGTGTCCGTATA 960
 Qy 961 AGCTTTTCTTCTGTGATCAAGGCTGCTGAGAGCCGCTGTAGAGATGAGCAGCAG 1020
 Db 961 AGCTTTTCTTCTGTGATCAAGGCTGCTGAGAGCCGCTGTAGAGATGAGCAGCAG 1020
 Qy 1021 CCAGAGAGAAATCGCTGTGACATTTGAGATTAAGATGAGAGCGGAGAACTTCAGCGAGC 1080
 Db 1021 CCAGAGAGAAATCGCTGTGACATTTGAGATTAAGATGAGAGCGGAGAACTTCAGCGAGC 1080
 Qy 1081 AGTGTGAGCTTGAGAGAGCGCCCAAGCGCTCTTGAGACAGACCCCAAGACAGAG 1140
 Db 1081 AGTGTGAGCTTGAGAGAGCGCCCAAGCGCTCTTGAGAGACAGACCCCAAGACAGAG 1140
 Qy 1141 CGTTGGCTCAGCTGAGCGCGCCGAGAGAGAGAGAGCGGAGCGCCAGAGACAG 1200
 Db 1141 CGTTGGCTCAGCTGAGCGCGCCGAGAGAGAGAGAGCGGAGCGCCAGAGACAG 1200
 Qy 1201 GAGGCGAAGCGGAGCTGAGCTGAGAGAGCTGAGAGAGCGGAGCTGAGAGCGG 1260
 Db 1201 GAGGCGAAGCGGAGCTGAGCTGAGAGAGCTGAGAGAGCGGAGCTGAGAGCGG 1260
 Qy 1261 CAGCGAG 1320
 Db 1261 CAGCGAG 1320
 Qy 1321 GAAAGGAG 1380
 Db 1321 GAAAGGAG 1380
 Qy 1381 AAAG 1440
 Db 1381 AAAG 1440
 Qy 1441 TTAGAGAGCTGAG 1500
 Db 1441 TTAGAGAGCTGAG 1500
 Qy 1501 CGAGTGGAG 1560
 Db 1501 CGAGTGGAG 1560
 Qy 1561 GGTGAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 Db 1561 GGTGAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 Qy 1621 ATTCCAG 1680
 Db 1621 ATTCCAG 1680

[illegible]

QY	2761	AAAGCAACCACTTAAATTTTAAACAAAGAGCATCATACCGTCTGGACAGCAAC	28230
Db	2761	AAAGCAACCACTTAAATTTTAAACAAAGAGCATCATACCGTCTGGACAGCAAC	28230
QY	2821	ATGTGTGTGTTTGGAGAACTTCAAGGTCAAGAGGTTGTTCCCAAGTCTTACGTAAA	28880
Db	2821	ATGTGTGTGTTTGGAGAACTTCAAGGTCAAGAGGTTGTTCCCAAGTCTTACGTAAA	28880
QY	2881	CTCATTTTCAGGGCCCCGTAAAGAAATCCACAGCATGATGATCTGAGGCTCTATGAAGTCT	29400
Db	2881	CTCATTTTCAGGGCCCCGTAAAGAAATCCACAGCATGATGATCTGAGGCTCTATGAAGTCT	29400
QY	2941	GCTAGTCTAAAGAGAGTGAGCTTCCCGGCCGCAAGCCAGCCATTCCGGAGAAAGTTT	30000
Db	2941	GCTAGTCTAAAGAGAGTGAGCTTCCCGGCCGCAAGCCAGCCATTCCGGAGAAAGTTT	30000
QY	3001	ATTGCCATGTACATATCAGAGACTTCTGACCAAGAGATTTTAACTTTACAGCAAGGAT	30660
Db	3001	ATTGCCATGTACATATCAGAGACTTCTGACCAAGAGATTTTAACTTTACAGCAAGGAT	30660
QY	3061	GTGATTTGTGTTTACCAAGAAAGATGTGTGATCTGTGTGACGGAAACGGTGGGCGACAAATCC	31260
Db	3061	GTGATTTGTGTTTACCAAGAAAGATGTGTGATCTGTGTGACGGAAACGGTGGGCGACAAATCC	31260
QY	3121	GGAGTCTTCCCTTCTAACTATGTGAGGCTTTAAAGATTCAAGAGGCTCTGAACTGTGGG	31800
Db	3121	GGAGTCTTCCCTTCTAACTATGTGAGGCTTTAAAGATTCAAGAGGCTCTGAACTGTGGG	31800
QY	3181	AAAACAGGAGTTTAAAGAAAAAAACCTGAAATTGGCCAGATTATGCTTCTGAGCTGCT	32400
Db	3181	AAAACAGGAGTTTAAAGAAAAAAACCTGAAATTGGCCAGATTATGCTTCTGAGCTGCT	32400
QY	3241	ACTGATCCCGAACCACTCACCTTGCTCTGTGGCAGCTGATTTCTGATCCGAAAAAGAAC	33000
Db	3241	ACTGATCCCGAACCACTCACCTTGCTCTGTGGCAGCTGATTTCTGATCCGAAAAAGAAC	33000
QY	3301	CCAGTGTGATGTGTGGGAAGGAAACTCGCAAGCTCCAGGGAAAAACCGCAGATTAAGGTGG	33660
Db	3301	CCAGTGTGATGTGTGGGAAGGAAACTCGCAAGCTCCAGGGAAAAACCGCAGATTAAGGTGG	33660
QY	3361	TTTTCACAGCAATTATGTCAAACTTCTTAACCTCCCGGAAACAGCAAAATCACCCCACTAG	34260
Db	3361	TTTTCACAGCAATTATGTCAAACTTCTTAACCTCCCGGAAACAGCAAAATCACCCCACTAG	34260
QY	3421	CTTACCCAAAGACCGCAGTGCAGCCAGCAGTGTGCAGGTGATCGGATGTACATTAAC	34800
Db	3421	CTTACCCAAAGACCGCAGTGCAGCCAGCAGTGTGTGCAGGTGATGTACATTAAC	34800
QY	3481	GCCAGAGCAGATGACGAACATGAGCTTACCAAAAGGCGAGATCATCAAGTCTCTCAACAG	35400
Db	3481	GCCAGAGCAGATGACGAACATGAGCTTACCAAAAGGCGAGATCATCAAGTCTCTCAACAG	35400
QY	3541	GAGGACCCGAGACTGTGTGAAAGGAGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCCATCAAT	36000
Db	3541	GAGGACCCGAGACTGTGTGAAAGGAGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCCATCAAT	36000
QY	3601	TATGTAAAGTGAACCAACAGATGAGCCCCACAGCAATGA	3642
Db	3601	TATGTAAAGTGAACCAACAGATGAGCCCCACAGCAATGA	3642

RESULT 2	
AAZ39025	
ID	AAZ39025 standard; cDNA; 5144 BP.
XX	
AC	AAZ39025;
XX	
DT	28-FEB-2000 (first entry)
XX	
DE	Mouse Eesl coding sequence.
XX	
KW	Mouse; murine; Eesl; Ees2; endocytosis; vesicular trafficking; regulation; actin cytoskeleton; detection; cancer; infection;

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 OS Mus sp.
 PN W09955728-A2.
 PD 04-NOV-1999.
 PF 27-APR-1999; 99MO-CA00375.
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX (HSCC-) HSC RES & DEV LP.
 PI Egan SE, Wang W, Sengar A;
 DR WPI; 2000-052802/04.
 DR P-PSDB; AAY57449.
 XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 XX
 PS Claim 6; Page 59-62; 99pp; English.
 XX
 CC The present invention specifically describes mammalian Ese1 and 2
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Eps15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese1 coding sequence.
 XX
 SO Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 other;
 Query Match 99.9%; Score 3639.4; DB 21; Length 5144;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 241 GCCATGAAGCTTATCAAACTGAAGCTAAGAGATATACGCTCCCTCCACACTTCCCTCCT 300
 Qy 301 GTCATGAAGAGCAACCAAGTGGCTATTTCCAGTGCACACAGCATTTTGTATAGAGGGATT 360
 Db 301 GTCATGAAGAGCAACCAAGTGGCTATTTCCAGTGCACACAGCATTTTGTATAGAGGGATT 360
 Qy 361 GCTAGCATGCCACACTCAACAGCTGTTCCTCTGTGCGCAATGGGCTCCCATTCAGTTGTT 420
 Db 361 GCTAGCATGCCACCACTCAACAGCTGTTCCTCTGTGCGCAATGGGCTCCCATTCAGTTGTT 420
 Qy 421 GGAATGTCTCAACCTAGTATCTTCTGTCCCTCCAGCAGCAGATGGCTCCCTGGCTAAC 480
 Db 421 GGAATGTCTCAACCTAGTATCTTCTGTCCCTCCAGCAGCAGATGGCTCCCTGGCTAAC 480
 Qy 481 GGGGCTCTCCCGTCATACAGCTTGCCTGCGTTTGGCATCTTGACGCACATGGCCA 540
 Db 481 GGGGCTCTCCCGTCATACAGCTTGCCTGCGTTTGGCATCTTGACGCACATGGCCA 540
 Qy 541 AAGAGTCTCTCCCTCAGCAGATCTGGTCCAGGGTCACTAATTAACAATAAGTACAGAG 600
 Db 541 AAGAGTCTCTCCCTCAGCAGATCTGGTCCAGGGTCACTAATTAACAATAAGTACAGAG 600
 Qy 601 GCACATCATTCGATGTGCGCAGGCCCCCTCCAGCAGCAGATGGGCTGTGCTCAGTCA 660
 Db 601 GCACATCATTCGATGTGCGCAGGCCCCCTCCAGCAGCAGATGGGCTGTGCTCAGTCA 660
 Qy 661 TCAAGCTGAATACAGGCGCTTATTTCAACAGCCAGACAAACTATGATGACACTTA 720
 Db 661 TCAAGCTGAATACAGGCGCTTATTTCAACAGCCAGACAAACTATGATGACACTTA 720
 Qy 721 ACAGGCTCCCGAGCAAGATTTTCAAGAGCCACACAAAATTAAGTGAAGCACTTA 780
 Db 721 ACAGGCTCCCGAGCAAGATTTTCAAGAGCCACACAAAATTAAGTGAAGCACTTA 780
 Qy 781 TCAATATGAATCTTTTGTGATGATCAATGAATGAAGAACTCACTGCAAGAAATTTATC 840
 Db 781 TCAATATGAATCTTTTGTGATGATCAATGAATGAAGAACTCACTGCAAGAAATTTATC 840
 Qy 841 CTAGCTATGACCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 841 CTAGCTATGACCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Qy 901 CCAGATATCATCT 960
 Db 901 CCAGATATCATCT 960
 Qy 961 AGCTCTTCTTCTGTGATCAAGAGCTCTGAGAGCGCTGTGAGAGATGAGACAG 1020
 Db 961 AGCTCTTCTTCTGTGATCAAGAGCTCTGAGAGCGCTGTGAGAGATGAGACAG 1020
 Qy 1021 CCAAGAGAAATCTGCTGTGATCAATTTGAAGATTAAGAGCGGGAATCTTGAAGAGGC 1080
 Db 1021 CCAAGAGAAATCTGCTGTGATCAATTTGAAGATTAAGAGCGGGAATCTTGAAGAGGC 1080
 Qy 1081 AGTGTGAGCTGTGAGAGAGCGCGCAAGCGCTCTTGAAGAGCAGAGCCCAAGACAGAG 1140
 Db 1081 AGTGTGAGCTGTGAGAGAGCGCGCAAGCGCTCTTGAAGAGCAGAGCCCAAGACAGAG 1140
 Qy 1141 CGGTGTGCTCAGCTGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 Db 1141 CGGTGTGCTCAGCTGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 Qy 1201 GAGGCGAAGCGGAGCTGAG 1260
 Db 1201 GAGGCGAAGCGGAGCTGAG 1260
 Qy 1261 CAGCGAG 1320
 Db 1261 CAGCGAG 1320
 Qy 1321 GAAAGGAG 1380
 Db 1321 GAAAGGAG 1380

```
QY 1381 AACAGGAGGAGGACCGTGTCTTGAAGCAAGAGAACTCTGAGTTTGAG 1440
|
|
|
Db 1381 AACAGGAGGAGGAGGACCGTGTCTTGAAGCAAGAGAACTCTGAGTTTGAG 1440
QY 1441 TTGGAAGCTGTAATGCAAAAAAGCATCAGCTGAAGGAAAACTTGAGATACAGGTGT 1500
|
|
|
Db 1441 TTGGAAGCTGTAATGCAAAAAAGCATCAGCTGAAGGAAAACTTGAGATACAGGTGT 1500
QY 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGAACTGAGAGAGCTTAAGAAAT 1560
|
|
|
Db 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGAACTGAGAGAGCTTAAGAAAT 1560
QY 1561 GCTGAATATCAACCACTTACAGCAGCAGTTGCAAGAAATCTCAGCAAAATGCTTGAGAGATT 1620
|
|
|
Db 1561 GCTGAATATCAACCACTTACAGCAGCAGTTGCAAGAAATCTCAGCAAAATGCTTGAGAGATT 1620
QY 1621 ATTCCAGAGAAACAGATACATCAGTACAGTAAACAACTCAGAGAAACATTTTGAT 1680
|
|
|
Db 1621 ATTCCAGAGAAACAGATACATCAGTACAGTAAACAACTCAGAGAAACATTTTGAT 1680
QY 1681 AGAGACTCGCTTTCTTACCTTCAAAAAGAGCTTGAAAGCAAGAGAGCTGGCCGAGCAG 1740
|
|
|
Db 1681 AGAGACTCGCTTTCTTACCTTCAAAAAGAGCTTGAAAGCAAGAGAGCTGGCCGAGCAG 1740
QY 1741 CTCGAGAGAGCTGAGCAGAGGTGAGAGAGAGACAGGTCAAAAGCTGAGAGATTTGAT 1800
|
|
|
Db 1741 CTCGAGAGAGCTGAGCAGAGGTGAGAGAGAGACAGGTCAAAAGCTGAGAGATTTGAT 1800
QY 1801 GTTTTCAACAACCAAGCTGAAGAACTGAGAGATACATGCAAAACAGCAATCCCAAG 1860
|
|
|
Db 1801 GTTTTCAACAACCAAGCTGAAGAACTGAGAGATACATGCAAAACAGCAATCCCAAG 1860
QY 1861 CAGAGGCTCTGAGAGGAGCGGCACTGAACAGAAAAGAGAGAGAGAGAGAGCTGAGAG 1920
|
|
|
Db 1861 CAGAGGCTCTGAGAGGAGCGGCACTGAACAGAAAAGAGAGAGAGAGAGAGCTGAGAG 1920
QY 1921 TTGAGAAAGCAAAAGAGAGAGCGCTCAGAGAGAGATTCAAGAAAGGAGCAAGCAATGGCTG 1980
|
|
|
Db 1921 TTGAGAAAGCAAAAGAGAGAGCGCTCAGAGAGAGATTCAAGAAAGGAGCAAGCAATGGCTG 1980
QY 1981 GAGCATGTGAGCAGAGAGAGAGCAGCCGCCCGGAAACCCCAAGAGAGAGAGAGAGAGAG 2040
|
|
|
Db 1981 GAGCATGTGAGCAGAGAGAGAGCAGCCGCCCGGAAACCCCAAGAGAGAGAGAGAGAGAG 2040
QY 2041 AAGAGGGAAGACAGTGTCAAGAAAGAAAGAGAGCGGAAAGAGAGAGAGAGAGAGAGAGAG 2100
|
|
|
Db 2041 AAGAGGGAAGACAGTGTCAAGAAAGAAAGAGAGCGGAAAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 GACAGAGAGAGTGGGCTTTTCCATCCGCAATCAGAGAGAGCTTAAGCTGGGCAACCCAGGCA 2160
|
|
|
Db 2101 GACAGAGAGAGTGGGCTTTTCCATCCGCAATCAGAGAGAGCTTAAGCTGGGCAACCCAGGCA 2160
QY 2161 CCGTGTCTTACCAAGAGAAAGCGGCTTACATTTCTGCAAGAGAGAGTGTAAAGTG 2220
|
|
|
Db 2161 CCGTGTCTTACCAAGAGAAAGCGGCTTACATTTCTGCAAGAGAGAGTGTAAAGTG 2220
QY 2221 GTATATTACCGAGCGGTGTACCCCTTTGAATCCAGAAAGTACAGTAGATCACCATCCAG 2280
|
|
|
Db 2221 GTATATTACCGAGCGGTGTACCCCTTTGAATCCAGAAAGTACAGTAGATCACCATCCAG 2280
QY 2281 CCAGAGATATAGTCAATGATGTAAGAGCAGACTGAGAGCAGAGATGAGCTTGAGAGA 2340
|
|
|
Db 2281 CCAGAGATATATAGTCAATGATGTAAGAGCAGACTGAGAGCAGAGATGAGCTTGAGAGA 2340
QY 2341 GAGCTGAAGAGGAAAGAGAGAGTGGTTCCTGCAAACTATGCAAGAAAGATTTCCAGAAAT 2400
|
|
|
Db 2341 GAGCTGAAGAGGAAAGAGAGAGTGGTTCCTGCAAACTATGCAAGAAAGATTTCCAGAAAT 2400
QY 2401 GAGGTTCCCACTCAGCAAAACAGATGACGATCTGACATCTGCCCCCTGCCCCAAATG 2460
|
|
|
Db 2401 GAGGTTCCCACTCAGCAAAACAGATGACGATCTGACATCTGCCCCCTGCCCCAAATG 2460

QY 2461 GCTCTGCTGAGACCCCTGCTCTTTGCAAGTGAAGCTTTCTGAGCCCTCCACAGACCCC 2520
|
|
|
Db 2461 GCTCTGCTGAGACCCCTGCTCTTTGCAAGTGAAGCTTTCTGAGCCCTCCACAGACCCC 2520
QY 2521 AACCACTGGGAGAGCTTCACTTCCAGCGTGGCCAGAGCTCAACAGAGAGAGAGAGAGAG 2580
|
|
|
Db 2521 AACCACTGGGAGAGCTTCACTTCCAGCGTGGCCAGAGCTCAACAGAGAGAGAGAGAGAG 2580
QY 2581 GACAACTGGGATACGTTGGGCGGCTCAGCTTTCTGACCGTACCTAGTGTGCTGGCAGTTA 2640
|
|
|
Db 2581 GACAACTGGGATACGTTGGGCGGCTCAGCTTTCTGACCGTACCTAGTGTGCTGGCAGTTA 2640
QY 2641 CGGCAAGATCAGCTTTTACCCAGCCAGCAGCTGCTCTCCCATCTCCGCTCTG 2700
|
|
|
Db 2641 CGGCAAGATCAGCTTTTACCCAGCCAGCAGCTGCTCTCCCATCTCCGCTCTG 2700
QY 2701 GGGCAGGTTGAAAAGGTGGAAGGGCTAACAGGAGCGGCTGATCTCCCTGAGAGCCAA 2760
|
|
|
Db 2701 GGGCAGGTTGAAAAGGTGGAAGGGCTAACAGGAGCGGCTGATCTCCCTGAGAGCCAA 2760
QY 2761 AAAGCAACCACTTAAATTTTAACAAAAGTGAAGCTCATCAAGCTTTGGAACAGCAAGAC 2820
|
|
|
Db 2761 AAAGCAACCACTTAAATTTTAACAAAAGTGAAGCTCATCAAGCTTTGGAACAGCAAGAC 2820
QY 2821 ATGTGTGTTTGGAGAAATTCAAGGTGAGAGAGGTTGTTCCCAAGTCTTACGTGAAA 2880
|
|
|
Db 2821 ATGTGTGTTTGGAGAAATTCAAGGTGAGAGAGGTTGTTCCCAAGTCTTACGTGAAA 2880
QY 2881 CTCATTTAGGGCCCGTGAAGAAATCCACAGCATCATCTGAGCCCTACGAAAGTCT 2940
|
|
|
Db 2881 CTCATTTAGGGCCCGTGAAGAAATCCACAGCATCATCTGAGCCCTACGAAAGTCT 2940
QY 2941 GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCAGAGCAGCATCTCCGAGAGAGATT 3000
|
|
|
Db 2941 GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCAGAGCAGCATCTCCGAGAGAGATT 3000
QY 3001 ATTGCATGTACACATACAGAGAGTTCTGAGCAGAGAGATTTAACCTTTCAAGAGGAGAT 3060
|
|
|
Db 3001 ATTGCATGTACACATACAGAGAGTTCTGAGCAGAGAGATTTAACCTTTCAAGAGGAGAT 3060
QY 3061 GTGATTTGTGTACCAAAAGATGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGATCC 3120
|
|
|
Db 3061 GTGATTTGTGTACCAAAAGATGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGATCC 3120
QY 3121 GAGGCTTCCCTTCTAATATGTAGAGGCTTAAAGATTCAAGAGGCTCTGAACTGTGGG 3180
|
|
|
Db 3121 GAGGCTTCCCTTCTAATATGTAGAGGCTTAAAGATTCAAGAGGCTCTGAACTGTGGG 3180
QY 3181 AAAACAGGAGTTTAAAGAAAAAACTGAAAAATGCGAGTTATGCTTCTACGCTGT 3240
|
|
|
Db 3181 AAAACAGGAGTTTAAAGAAAAAACTGAAAAATGCGAGTTATGCTTCTACGCTGT 3240
QY 3241 ACTGTCTCCGAAACAACTACCTGCTCTGAGCAGCTGATTTGATTCGGAAAAAGAAC 3300
|
|
|
Db 3241 ACTGTCTCCGAAACAACTACCTGCTCTGAGCAGCTGATTTGATTCGGAAAAAGAAC 3300
QY 3301 CCAGGTGATGTGGGAAAGAGAACTGCAAGCTGAGGAGAAAAAGCGCAGATGAGGGTG 3360
|
|
|
Db 3301 CCAGGTGATGTGGGAAAGAGAACTGCAAGCTGAGGAGAAAAAGCGCAGATGAGGGTG 3360
QY 3361 TTTCCAGCAAAATTAATGTCAACTTAAAGCCCGGAAACAGCAAAATCAACCCAACTGAG 3420
|
|
|
Db 3361 TTTCCAGCAAAATTAATGTCAACTTAAAGCCCGGAAACAGCAAAATCAACCCAACTGAG 3420
QY 3421 CTACCAAGAGCCGAGTGCAGCAGCAGTGTGCAAGTGTGCGGATGTACAGATTAAC 3480
|
|
|
Db 3421 CTACCAAGAGCCGAGTGCAGCAGCAGTGTGCAAGTGTGCGGATGTACAGATTAAC 3480
QY 3481 GCCCAGAGCAGTGAAGAACTAGCTTCAAGAAAGCCAGATCATCAAGTCTCAAG 3540
|
|
|
Db 3481 GCCCAGAGCAGTGAAGAACTAGCTTCAAGAAAGCCAGATCATCAAGTCTCAAG 3540
QY 3541 GAGGACCCGAGCTGTGTGAAAAGAGAGTCAAGTGGGCAAGTTGGGCTTTCCATCCAAAT 3600
|
|
|
```


QY 1081 AGTGTGAGCTGAGAAAGCCGCGCAAGCCGCTTGTGAGCAGCAGCGCAAGAGCAGAG 1140
DB 1340 AGTGTGAGCTGAGAAAGCCGCGCAAGCCGCTTGTGAGCAGCAGCGCAAGAGCAGAG 1399
QY 1141 CGGTTGGCTCAGCTGAGACCGCGCGCAGCAGAGAGAGAAAGAGCGGGAGCGCCAGAGCAG 1200
DB 1400 CGGTTGGCTCAGCTGAGACCGCGCGCAGCAGAGAGAGAGAGCGGGAGCGCCAGAGCAG 1459
QY 1201 GAGGCCAAGCGCAGCTGAGAGTGTGAGAAAGCAGCTGAGAGCAGCGGAGCTGAGACCG 1260
DB 1460 GAGGCCAAGCGCAGCTGAGAGTGTGAGAAAGCAGCTGAGAGCAGCGGAGCTGAGACCG 1519
QY 1261 CACCGAGAGAGAGAGAGAGAGAGATGAGAGCGCGAGCGCGCAAAACCGGAACTG 1320
DB 1520 CACCGAGAGAGAGAGAGAGAGATGAGAGCGCGAGCGCGCAAAACCGGAACTG 1579
QY 1321 GAAAGGACGAGAACTTGAATGGGAAACCGAGACAGGAACTCTGTAATCAGAG 1380
DB 1580 GAAAGGACGAGAACTTGAATGGGAAACCGAGACAGGAACTCTGTAATCAGAG 1639
QY 1381 AACAAAG 1440
DB 1640 AACAAAG 1699
QY 1441 TTGAAGCTCTGAATGACAAAAGACATCAGCTGAGAGAGAACTTTCAGAGATCAGGTGT 1500
DB 1700 TTGAAGCTCTGAATGACAAAAGACATCAGCTGAGAGAGAACTTTCAGAGATCAGGTGT 1759
QY 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGACAGAACTGTAAGAGCTAAGAAATT 1560
DB 1760 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGACAGAACTGTAAGAGCTAAGAAATT 1819
QY 1561 GCTGAATCACTCCACTTACAGCAGAGTTGACAGAACTCAGCAAAATGCTGGAAGACTT 1620
DB 1820 GCTGAATCACTCCACTTACAGCAGAGTTGACAGAACTCAGCAAAATGCTGGAAGACTT 1879
QY 1621 ATTCCAGAGAAACAGATACTCAGTGAACAGTTTAAACAAAGTCAGAGAACTCCAGAA 1680
DB 1880 ATTCCAGAGAAACAGATACTCAGTGAACAGTTTAAACAAAGTCAGAGAACTCCAGAA 1939
QY 1681 AGAGACTCGCTTTTACCTTCAAAAGAGCTTGAAGCAAAAGAGCTGGCGCAGCAG 1740
DB 1940 AGAGACTCGCTTTTACCTTCAAAAGAGCTTGAAGCAAAAGAGCTGGCGCAGCAG 1999
QY 1741 CTCGGAGAGAGCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1800
DB 2000 CTCGGAGAGAGCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2059
QY 1801 GTTTTCAACAAACAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGCACTCCAGAA 1860
DB 2060 GTTTTCAACAAACAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGCACTCCAGAA 2119
QY 1861 CAGAGTCCCTGAGAGAGAGCTGAAAGCAGAAAGAGAGAGAGAGAGAGAGAGCTGAG 1920
DB 2120 CAGAGTCCCTGAGAGAGAGCTGAAAGCAGAAAGAGAGAGAGAGAGAGAGAGCTGAG 2179
QY 1921 TTGAGAGAGAGAGAGAGAGAGCTCAGAGAGAGTTCAGAGAAAGAGAGAGAGAGAGAG 1980
DB 2180 TTGAGAGAGAGAGAGAGAGAGCTCAGAGAGAGTTCAGAGAAAGAGAGAGAGAGAGAG 2239
QY 1981 GAGCATGTGAG 2040
DB 2240 GAGCATGTGAG 2299
QY 2041 AAGAGGAG 2100
DB 2300 AAGAGGAG 2359
QY 2101 GACAG 2160
DB 2360 GACAG 2419

QY 2161 CCTGTCTACACAGAGAAAGCCCGCTTACATTTCTGACAGAGAGTGTAAAGTG 2220
DB 2420 CCTGTCTACACAGAGAAAGCCCGCTTACATTTCTGACAGAGAGTGTAAAGTG 2479
QY 2221 GTAATTTACGAGGCGCTGATCCCTTTGAATCCGAGAGTACGATGAGATCAGATCAG 2280
DB 2480 GTAATTTACGAGGCGCTGATCCCTTTGAATCCGAGAGTACGATGAGATCAGATCAG 2539
QY 2281 CCAGAGATATAGTCAATGTGATGAAAGCAGACTGAGAGAGCAGAGATGCTTGAAGA 2340
DB 2540 CCAGAGATATAGTCAATGTGATGAAAGCAGACTGAGAGAGCAGAGATGCTTGAAGA 2599
QY 2341 GAGCTGAAAGGAG 2400
DB 2600 GAGCTGAAAGGAG 2659
QY 2401 GAGGTTCCACTCCAGCAAAACAGTACCGATGATGATCTGCCCCCTGCCAACTG 2460
DB 2660 GAGGTTCCACTCCAGCAAAACAGTACCGATGATGATCTGCCCCCTGCCAACTG 2719
QY 2461 GCTCTGAG 2520
DB 2720 GCTCTGAG 2779
QY 2521 AACAACTGGGAGAGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
DB 2780 AACAACTGGGAGAGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2839
QY 2581 GACAACTGGGATATAGTGGGCGGCTCAGCTTCTTGACCGTACCTAGTGTGCGCACTTA 2640
DB 2840 GACAACTGGGATATAGTGGGCGGCTCAGCTTCTTGACCGTACCTAGTGTGCGCACTTA 2899
QY 2641 CGGAGAGATCAGCTTTTACCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
DB 2900 CGGAGAGATCAGCTTTTACCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2959
QY 2701 GCGCAGAGTGAAG 2760
DB 2960 GCGCAGAGTGAAG 3019
QY 2761 AAAGACAACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
DB 3020 AAAGACAACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3079
QY 2821 ATGTGTGTGTTGAGAGAGAGTCAAGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
DB 3080 ATGTGTGTGTTGAGAGAGAGTCAAGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3139
QY 2881 CTCATTTCAAGGCGCGTAAAGAAATCCAAAGCATGATCTGAGCTTACGAAAGTCTT 2940
DB 3140 CTCATTTCAAGGCGCGTAAAGAAATCCAAAGCATGATCTGAGCTTACGAAAGTCTT 3199
QY 2941 GCTAGTCTTAAAGAGAGAGTCTTCCCGCGCGCAAGCAGCATTTCCCGAGAGAGAGTT 3000
DB 3200 GCTAGTCTTAAAGAGAGAGTCTTCCCGCGCGCAAGCAGCATTTCCCGAGAGAGAGTT 3259
QY 3001 ATTGCATGTACATACGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
DB 3260 ATTGCATGTACATACGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3319
QY 3061 GTGATTGTGTTTAAAGAAAGATGTGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
DB 3320 GTGATTGTGTTTAAAGAAAGATGTGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3379
QY 3121 GGAAGTCTTCCCTTAACTATGTGAGGCTTAAAGATCAAGAGGCTTGTGAATGCTGCGG 3180
DB 3380 GGAAGTCTTCCCTTAACTATGTGAGGCTTAAAGATCAAGAGGCTTGTGAATGCTGCGG 3439
QY 3181 AAAACAGGAGTTTAAAG 3240
DB 3440 AAAACAGGAGTTTAAAG 3499
QY 3241 ACTGGTCCGAACAACTCACCTGTCTCTGGGAGAGCTGATTTGATCCGAGAAAGAAC 3300

Db 3500 ACTGATCCGAACTACCTCCTGGAGTGTCTGATCCGAAAAAGAAC 3559
 QY 3301 CCAAGTGTATGTGGGAAAGAACTTCAGCTCCAGGAAAAAGCCAGATAGGCTG 3360
 Db 3560 CCAAGTGTATGTGGGAAAGAACTTCAGCTCCAGGAAAAAGCCAGATAGGCTG 3619
 QY 3361 TTTCCAGCAATATGTCAAACTTCTAAGCCCGGAAACAAGAAATCAACCCCACTGAG 3420
 Db 3620 TTTCCAGCAATATGTCAAACTTCTAAGCCCGGAAACAAGAAATCAACCCCACTGAG 3679
 QY 3421 CTAACCAAGACCCGAGTGCAGCCAGCAGTGTCCAGTGTATGGGATGTCATTCACCC 3480
 Db 3680 CTAACCAAGACCCGAGTGCAGCCAGCAGTGTGTCCAGTGTATGGGATGTCATTCACCC 3739
 QY 3481 GCCCAGAACGATGACGAACTAGCTTCCAGCAAGGCGAGATCATCAACCTCCACCAAG 3540
 Db 3740 GCCCAGAACGATGACGAACTAGCTTCCAGCAAGGCGAGATCATCAACCTCCACCAAG 3799
 QY 3541 GAGGACCCGAGTGTGGGAAAGAAAGTCAAGTGGGCAAGTTGGGCTTTCCCATCAAT 3600
 Db 3800 GAGGACCCGAGTGTGGGAAAGAAAGTCAAGTGGGCAAGTTGGGCTTTCCCATCAAT 3859
 QY 3601 TATGTAAGCTGACACACAGACATGAGACCCAGCCAGCAATG 3641
 Db 3860 TATGTAAGCTGACACACAGACATGAGACCCAGCCAGCAATG 3900

RESULT 4

AAZ39008

ID AAZ39008 standard; cDNA; 5082 BP.

AAZ39008;

XX 28-FEB-2000 (first entry)

DE Mouse Esei full length cDNA sequence.

XX Mouse; murine; Esei; Esei2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KM EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KM antiproliferative; antiviral; ss.

XX Mus sp.

PN W0955728-A2.

XX 04-NOV-1999.

PF 27-APR-1999; 99MO-CA00375.

PR 27-APR-1998; 98CA-2230201.

PR 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

PI Egan SE, Wang W, Sengar A;

XX WPI; 2000-052802/04.

DR P-PSDB; AAY57444.

XX New nucleic acid encoding Esei and 2 proteins, involved in regulation

PT of endocytosis, used e.g. for treating cancer or preventing viral

PT infection -

XX Claim 6; Page 38-40; 99pp; English.

XX The present sequence encodes mouse Esei. The present invention
 CC specifically describes mammalian Esei and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Epsi1 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive

CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Esei genes or
 CC antagonists of an Esei binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Esei is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Esei mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Esei-Epsi1 complex, then binding dynamn to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC XX

SQ Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 other;

Query Match 99.7%; Score 3630; DB 21; Length 5082;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;

Matches 3641; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCCTGGATGTCGGCCATPACTGTGAG 60

Db 259 ATGGCTCAGTTTCCACACCTTTGGTGTAGCCTGGATGTCGGCCATPACTGTGAG 318

QY 61 GAAAGGCGCAAGCATGACAGCAGTCTTACCTGAAAGCCGATAGCGGATTTATTA 120

Db 319 GAAAGGCGCAAGCATGACAGCAGTCTTACCTGAAAGCCGATAGCGGATTTATTA 378

QY 121 GGTGATCAAGGAGGAACTTTTTCATCTGGTATCCTGAGCTGTCTTACGACAA 180

Db 379 GGTGATCAAGGAGGAACTTTTTCATCTGGTATCCTGAGCTGTCTTACGACAA 438

QY 181 ATATGGGCGCTAGCGGACATGATAGAGTGAAGATGATCAATGGAATTTTCCATA 240

Db 439 ATATGGGCGCTAGCGGACATGATAGAGTGAAGATGATCAATGGAATTTTCCATA 498

QY 241 GCCATGAAGCTTATCAAACTGAAGCTCAAGGATATGAGTCCCTCCACACTTCCCT 300

Db 499 GCCATGAAGCTTATCAAACTGAAGCTCAAGGATATGAGTCCCTCCACACTTCCCT 558

QY 301 GTCATGAAGCAACCAAGCTGATTTTCCAGTCCACAGCATTTGGTATAGAGGAT 360

Db 559 GTCATGAAGCAACCAAGCTGATTTTCCAGTCCACAGCATTTGGTATAGAGGAT 618

QY 361 GTTACATGCCACCACTACAGCTGTGCTCTCTGCAATGGGCTCATTCAGATT 420

Db 619 GTTACATGCCACCACTACAGCTGTGCTCTCTGCAATGGGCTCATTCAGATT 678

QY 421 GGAATGTCCCAACCTTATGATCTTCTGCTCCAGAGCAGTCCCTCCCTGGTAA 480

Db 679 GGAATGTCCCAACCTTATGATCTTCTGCTCCAGAGCAGTCCCTCCCTGGTAA 738

QY 481 GGGGCTCTCCGCTATACAGCTGTGCTGCTGTTGGCATCTTGCACCAATGCGCA 540

Db 739 GGGGCTCTCCGCTATACAGCTGTGCTGCTGTTGGCATCTTGCACCAATGCGCA 798

QY 541 AAGAATTTCTTCTTACAGAGATCTGTCCAGAGTCAAACTTAACACTTAAGTTACAGAG 600

Db 799 AAGAATTTCTTCTTACAGAGATCTGTCCAGAGTCAAACTTAACACTTAAGTTACAGAG 858

QY 601 GCACATCATTTGATGTGCCAGGCGCCCTCCAGAGCAGATGGGCTGTGCTCAGTCA 660

Db 859 GCACATCATTTGATGTGCCAGGCGCCCTCCAGAGCAGATGGGCTGTGCTCAGTCA 918

QY 661 TCAAGCTGAAATACAGCAGTATTTCAACGCCACCAAAACTATAGTGGACACTTA 720

Db 919 TCAAGCTGAAATACAGCAGTATTTCAACGCCACCAAAACTATAGTGGACACTTA 978

QY 721 AAGGATCCCGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780

Db 979 AAGGATCCCGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 1038

QY 781 TCAATATGAAATCTTTCTGACATTGATCAAGATGAAAACCTCACTGACAGAAATTTATC 840
 Db 1039 TCAATATGAAATCTTTCTGACATTGATCAAGATGAAAACCTCACTGACAGAAATTTATC 1098
 QY 841 CTAGCTATGACCAATATGTAATGTTGCCATGTCGTGACGACCACTGCGCCGCTCCGCT 900
 Db 1099 CTAGCTATGACCAATATGTAATGTTGCCATGTCGTGACGACCACTGCGCCGCTCCGCT 1158
 QY 901 CCAGAAATACATCCCTCCTCTCTCAGAAAGATTGCTCCGCGAGTGGAGTGTCCGTGATA 960
 Db 1159 CCAGAAATACATCCCTCCTCTCTCAGAAAGATTGCTCCGCGAGTGGAGTGTCCGTGATA 1218
 QY 961 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGCTCAGAGATGACGACAG 1020
 Db 1219 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGCTCAGAGATGACGACAG 1278
 QY 1021 CCAGAGAAAGAACTGCTGCTGATCATTGAAATAGAAAGCGGAGAACTTTCAGCGAGGC 1080
 Db 1279 CCAGAGAAAGAACTGCTGCTGATCATTGAAATAGAAAGCGGAGAACTTTCAGCGAGGC 1338
 QY 1081 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGAGCAGCGCAAGAGCAGAG 1140
 Db 1339 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGAGCAGCGCAAGAGCAGAG 1398
 QY 1141 CGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGGAGCGCCAGAGAG 1200
 Db 1399 CGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGGAGCGCCAGAGAG 1458
 QY 1201 GAGGCCAAGCGGAGCTGAGAGCTGAGAGAGCAGTGAAGAGCAGCGGAGCTGAGCGG 1260
 Db 1459 GAGGCCAAGCGGAGCTGAGAGAGCAGTGAAGAGCAGCGGAGCTGAGCGG 1518
 QY 1261 CAGCAGAGAGAGAGAGAGAGAGATGAGAGCGCGCAGCGCCCAAAACCGGAACTG 1320
 Db 1519 CAGCAGAGAGAGAGAGAGAGATGAGAGCGCGCAGCGCCCAAAACCGGAACTG 1578
 QY 1321 GAAAGCAGCAGCAACTTGAATGGAAACGGAACCGGAGCAGGAACCTCTGAATCAGAG 1380
 Db 1579 GAAAGCAGCAGCAACTTGAATGGAAACGGAACCGGAGCAGGAACCTCTGAATCAGAG 1638
 QY 1381 AACAAAG 1440
 Db 1639 AACAAAG 1698
 QY 1441 TTGAAAGCTCTGAATGACAAAAGAGATCAGCTAGAGAGAACTTCAAGATACAGTGT 1500
 Db 1699 TTGAAAGCTCTGAATGACAAAAGAGATCAGCTAGAGAGAACTTCAAGATACAGTGT 1758
 QY 1501 CGACTGGCAACCCAGAGAGCAAGAAATTTAGAGCAGCAAGCTTGAAGAGCTAAGATTT 1560
 Db 1759 CGACTGGCAACCCAGAGAGCAAGAAATTTAGAGCAGCAAGCTTGAAGAGCTAAGATTT 1818
 QY 1561 GCTGAATATCACCCTTACAGCAGCAGTGTGAGAGATTTTCAGCAAAATGCTTGAAGACTT 1620
 Db 1819 GCTGAATATCACCCTTACAGCAGCAGTGTGAGAGATTTTCAGCAAAATGCTTGAAGACTT 1878
 QY 1621 ATTCCAGAGAAACAGATATCAGTGAACAGTTTAAACAAATCCAGAGAAACAGTTTGCAT 1680
 Db 1879 ATTCCAGAGAAACAGATATCAGTGAACAGTTTAAACAAATCCAGAGAAACAGTTTGCAT 1938
 QY 1681 AGAGACTGCTTCTTACCTCAAAAGAGCCTTGAAGCAAAAGAGCTGCGCGGAGCAG 1740
 Db 1939 AGAGACTGCTTCTTACCTCAAAAGAGCCTTGAAGCAAAAGAGCTGCGCGGAGCAG 1998
 QY 1741 CTCGGGAGAGAGCTGAGAGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 Db 1999 CTCGGGAGAGAGCTGAGAGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2058
 QY 1801 GTTTTCAACAAACAGCTGAAGAGAACTGAGAGAGATCATAGCAAAACAGCAACTCCAGAG 1860
 Db 2059 GTTTTCAACAAACAGCTGAAGAGAACTGAGAGAGATCATAGCAAAACAGCAACTCCAGAG 2118

QY 1861 CAGAGTCCCTGAGAGCGCCGCACTGAAGCAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1920
 Db 2119 CAGAGTCCCTGAGAGCGCCGCACTGAAGCAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2178
 QY 1921 TTAGAGAGCAAAAG 1980
 Db 2179 TTAGAGAGCAAAAG 2238
 QY 1981 GAGCATGTGAG 2040
 Db 2239 GAGCATGTGAG 2298
 QY 2041 AAGAGAGAAACAGTGTGAG 2100
 Db 2299 AAGAGAGAAACAGTGTGAG 2358
 QY 2101 GACAAAGAGAGTCCGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 Db 2359 GACAAAGAGAGTCCGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2418
 QY 2161 CCTGCTTACACAGAGAAAGCGCCGCTTACATTTCTGACAGAGAGAGTAAAGTG 2220
 Db 2419 CCTGCTTACACAGAGAAAGCGCCGCTTACATTTCTGACAGAGAGAGTAAAGTG 2478
 QY 2221 GTATATTACGAGCGCTGTACCTCTTTGAATCCAGAGTACAGATGATCAGATCCATCAG 2280
 Db 2479 GTATATTACGAGCGCTGTACCTCTTTGAATCCAGAGTACAGATGATCAGATCCATCAG 2538
 QY 2281 CCAGAGATATAGTATAGTGTGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
 Db 2539 CCAGAGATATAGTATAGTGTGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2598
 QY 2341 GAGCTGAAG 2400
 Db 2599 GAGCTGAAG 2658
 QY 2401 GAGGTTCCCACTCCAGCAACCAAGTACAGATGATGATGATGATGATGATGATGATGATG 2460
 Db 2659 GAGGTTCCCACTCCAGCAACCAAGTACAGATGATGATGATGATGATGATGATGATGATG 2718
 QY 2461 GCTCTGCTGAG 2520
 Db 2719 GCTCTGCTGAG 2778
 QY 2521 AACAACTGGGAGAGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
 Db 2779 AACAACTGGGAGAGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2838
 QY 2581 GACAACTGGGATACGAG 2640
 Db 2839 GACAACT - GATATGAG 2697
 QY 2641 CCGCAGAGATCAGGCTTTTACCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
 Db 2898 CCGCAGAGATCAGGCTTTTACCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2957
 QY 2701 GAGCAGGAGTGAAG 2760
 Db 2958 GAGCAGGAGTGAAG 3017
 QY 2761 AAAGCAACCACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
 Db 3018 AAAGCAACCACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3077
 QY 2821 ATGTGTGTGTTTGAAG 2880
 Db 3078 ATGTGTGTGTTTGAAG 3137
 QY 2881 CTCATTTCAAGGAG 2940
 Db 3138 CTCATTTCAAGGAG 3197
 QY 2941 GCTAGTCTAAAG 3000

```

Db 3198 GCTAGCTTAAGAGAGTGGCTTCCCGGCCCAAGCCGACCTTCCCGAGAGAGTTT 3257
Qy 3001 ATTGGCATGTACATACGAGAGTTCTGAGCAAGAGATTAACTTTTACGAAAGGGAT 3060
Db 3258 ATTGGCATGTACATACGAGAGTTCTGAGCAAGAGATTAACTTTTACGAAAGGGAT 3317
Qy 3061 GTGATTGTGTTTCCAAAGAAAGATGTGACTGTGACGGGAAACGGTGGGCAACAAGTCC 3120
Db 3318 GTGATTGTGTTTCCAAAGAAAGATGTGACTGTGACGGGAAACGGTGGGCAACAAGTCC 3377
Qy 3121 GGAATCTTCCCTCTCTAATCTATGTGAGGCTTAAAGATTCAAGGGCTCTGGAATGCTGGG 3180
Db 3378 GGAATCTTCCCTCTCTAATCTATGTGAGGCTTAAAGATTCAAGGGCTCTGGAATGCTGGG 3437
Qy 3181 AAAACAGGGAGTTTAGGAAAAAAACCTGAAATTTGCCAGATTATGCTTCTCAAGCTGCT 3240
Db 3438 AAAACAGGGAGTTTAGGAAAAAAACCTGAAATTTGCCAGATTATGCTTCTCAAGCTGCT 3497
Qy 3241 ACTGATCCCGAACAACCTCACTGCTGCTGAGGAGCTGATTTCTGATCCGAAAAAAGAAC 3300
Db 3498 ACTGATCCCGAACAACCTCACTGCTGCTGAGGAGCTGATTTCTGATCCGAAAAAAGAAC 3557
Qy 3301 CCAAGTGTGATGTGGGAAAGAACTGCAAGCTCGAGGGAAAAAGGCCAGATAGGGTGG 3360
Db 3558 CCAAGTGTGATGTGGGAAAGAACTGCAAGCTCGAGGGAAAAAGGCCAGATAGGGTGG 3617
Qy 3361 TTTCCAGCAATATATGCAAACTTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3420
Db 3618 TTTCCAGCAATATATGCAAACTTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3677
Qy 3421 CTACCCAAAGCCGAGTGCAGCCAGAGTGTGCAAGTGTGCGGATGTACATTCACACC 3480
Db 3678 CTACCCAAAGCCGAGTGCAGCCAGAGTGTGCAAGTGTGCGGATGTACATTCACACC 3737
Qy 3481 GCCCAGAGCATGACGAACTAGCTTCAAGAAAGCCAGATCATCAACGCTCCCAACAAG 3540
Db 3738 GCCCAGAGCATGACGAACTAGCTTCAAGAAAGCCAGATCATCAACGCTCCCAACAAG 3797
Qy 3541 GAGGACCCGAGTGTGGAAAGGAAAGTCAAGTGGGCAAGTTGGGCTTTCCCATCAAT 3600
Db 3798 GAGGACCCGAGTGTGGAAAGGAAAGTCAAGTGGGCAAGTTGGGCTTTCCCATCAAT 3857
Qy 3601 TATGTAAAGCTGACCAACAGACATGAGACCCCAAGCCAGCAATGA 3642
Db 3858 TATGTAAAGCTGACCAACAGACATGAGACCCCAAGCCAGCAATGA 3899

RESULT 5
AAZ34572 standard; cDNA; 5195 BP.
ID AAZ34572;
AC AAZ34572;
DT 01-FEB-2000 (first entry)
DE Human SH3D1A cDNA clone 11.
XX
XX SH3D1A gene; human; Down's syndrome; leukemia; cancer;
XX megakaryocytic abnormality; myeloproliferative disorder;
XX platelet disorder; neural disorder; thrombocytopenia;
XX hemotopoietic disorder; cognitive dysfunction; microcephaly;
XX lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
XX ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 239..3886
XX FT CDS
XX FT /+tag= a
XX
XX W09953062-AA2.

```

```

PD 21-OCT-1999.
XX
XX 16-APR-1999; 99MO-US08371.
XX
XX 16-APR-1998; 98US-0082007.
XX
XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX
XX Korenberg JR, Chen X;
XX WPI: 1999-633829/54.
XX P-PSDB; AAY32156.
XX
XX Nucleic acid from the human SH3D1A gene and its products, useful for
XX the diagnosis and treatment of myeloproliferative disorders and
XX leukemia.
XX
XX Claim 2; Fig 10; 99pp; English.

```

This is the nucleotide sequence of full-length cDNA (clone 11) corresponding to a novel human SH3 gene, termed the SH3D1A gene, that contributes to the development of platelets and the pathogenesis of leukemias, both in general and in particular those involving the megakaryocytic lineage. The SH3D1A gene maps to the small candidate region for low platelets on chromosome 21. Sequencing of 5 different sizes of cDNA clone from foetal brain (see AAZ34570-74) suggests that at least 3 isoforms exist. The invention provides methods for the diagnosis and treatment of megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, acute leukemia, neural disorders, thrombocytopenia, platelet disorder on chromosome 21, low platelets in deletion for 21, association of gains in chromosome 21 with leukemias, neural abnormalities, dysfunctions and disorders including brain malformations and corresponding cognitive dysfunctions, microcephaly, lissencephaly, and colpocephaly. Methods are also provided for: suppressing cells unable to regulate themselves; screening for a somatic alteration in the SH3D1A gene; monitoring the progress and adequacy of a treatment; monitoring tumour risk progress or megakaryocytic abnormality; platelet disorder or leukemia; and haematopoietic disorder, platelet disorder or leukemia; and treatment of a subject (including a prenatal subject) having megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or neural disorder using a nucleic acid that expresses SH3D1A or its antisense nucleic acid.

Sequence 5195 BP; 1592 A; 1148 C; 1237 G; 1218 T; 0 other;

Query Match 78.4%; Score 2856; DB 20; Length 5195;
Best Local Similarity 86.8%; Pred. No. 0;
Matches 3167; Conservative 0; Mismatches 475; Indels 6; Gaps 2;

```

Qy 1 ATGGCTCAGTTTCCCAACACCTTTGGTGTGAGCTGATGCTGGGCATTAAGTGGAG 60
Db 239 ATGGCTCAGTTTCCCAACACCTTTGGTGTGAGCTGATGCTGGGCATTAAGTGGAG 298
Qy 61 GAAAGGGCCAAAGCATGACACAGCTTCTTAGCCTGAAGCCGATAGCGGATTTATTACT 120
Db 229 GAAAGGGCCAAAGCATGACACAGCTTCTTAGCCTGAAGCCGATAGCGGATTTATTACT 358
Qy 121 GGTGATCAAGGAGGAATTTTTTTTCCATCTGGGTAACTCTGACCTGCTTGAACAA 180
Db 359 GGTGATCAAGGAGGAATTTTTTTTCCATCTGGGTAACTCTGACCTGCTTGAACAA 418
Qy 181 ATATGGGCGCTAGCGGACATGAATACGATGGAAGATGATCAAGTGAATTTTCATA 240
Db 419 ATATGGGCGCTAGCGGACATGAATACGATGGAAGATGATCAAGTGAATTTTCATA 478
Qy 241 GCCATGAAGCTTATCAAACTGAAGCTACAGAGATATGACTCCCTCCACACTTCCCTCT 300
Db 479 GCTATGAAGCTTATCAAACTGAAGCTACAGAGATATGACTCCCTCCACACTTCCCTCT 538
Qy 301 GTCATGAAGCGAACCGAGTGTATTTCCAGTGCACAGCATTTGGATGTGAGGAGTT 360

```

Dh 539 GTCATAAAACAGCAACCACTGCTATTTCTAGGCAACACATTTGGTATGGAGGTATC 598
Qy 361 GCTAGCATGCCACACTCAACAGCTGTGCTCTGTGCCAATGGGCTCCATTCAGTTGTT 420
Dh 599 GCCAGATGCCACCGCTTACAGCTGTGCTCCAGTGCATGGGATCCATTCAGTTGTT 658
Qy 421 GGAATGTTCCACCTTACTATTTCTGTCCCTCCAGCAGCAGTGGCTCCCTGGCTAAC 480
Dh 659 GGAATGTTCCAAACCTAGATCTTGTGTCCCAAGCAGCTGTGGCCCCCTGGCTAAC 718
Qy 481 GGGGCTCCCTCCGCTATACAGCTCTGCTGGGCTTGGCATCTCCAGCACCATGGCCA 540
Dh 719 GGGGCTCCCTCTGTATTAACAACCTCTGCTGATTTGCTATCTCCAGCACCATGGCCA 778
Qy 541 AAGAGTTCTTCTTACAGCATCTGTGTCCAGGCTCAATTAACACTAAGTTACAGAA 600
Dh 779 AAGAGTTCTTCTTGTAGATCTGTGTCCAGGCTCAAACTAAACCTAATTTACAAAAG 838
Qy 601 GCACATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAGAAATGGCTGTGCTCACTCA 660
Dh 839 GCACATCATTTGATGTGCGCAGGTGCCACAGTGCGAGAGTGGGCTGTCTCCTCAGTCA 898
Qy 661 TCAGGCTCAAAATACAGGAGGATTTATCAACAGCAGCAAAACTATGATGACACTTA 720
Dh 899 TCAGGCTCAAAATACAGGCAATTTATTAATAGTCAAGCAAAACTATGATGACACTTA 958
Qy 721 ACAGGTCCTCCAGCAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 780
Dh 959 ACAGGTCCTCCAGCAGAACTATTTCTATGCAGTCAAGTTTACCAAGCTCAGCTGGCT 1018
Qy 781 TCATATGGAATCTTTCTGACATTTGATCAAGATGAAACTCATGCGAGAAAGATTTTATC 840
Dh 1019 TCATATGGAATCTTTCTGACATTTGATCAAGATGAAACTCATGCGAGAAATTTTATC 1078
Qy 841 CTAGCATGCACTTAATGATGTTGCTGATGCTGCTCAGCACTGGCGGCTGCTGCT 900
Dh 1079 CTAGCATGCACTTAATGATGTTGCTGATGCTGCTGCTCAGCACTGGCGGCTGCTGCT 1138
Qy 901 CCAGAAATACATCCCTCTCTTCTTACAGAAAGTTGCTCCGCGAGTGGATGTCCTCATTA 960
Dh 1139 CCAGAAATACATCCCTCTTCTTACAGAAAGTTGCTCCGCGAGTGGATGTCCTCATTA 1198
Qy 961 AGCTCTTCTTCTGTGATCAAGGCTGCTGAGAGCCGCTCGTCAAGATGACACAG 1020
Dh 1199 AGCTCAACATCTGTAGATCAAGGCTTACCAAGGAACTGTTTGAAGATGAACAACA 1258
Qy 1021 C---CAGAGAAAGAACTGCTGTGACATTTGAATTAAGGGGAGAACTTCCAGGCA 1077
Dh 1259 CAAATTGAAAAGAAATTTACCTGTAACTTTGAAGATTAAGGGGAGAACTTTGAAGCT 1318
Qy 1078 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCAAGAGCAG 1137
Dh 1319 GGCAGTGTGAGCTGAGAGAGCGAAAGCAAGCTCTCTGGAACAGAGCGCAAGAGCAG 1378
Qy 1138 GAGCGTTGGCTCAGCTGAGCGCGCGCAGCAGAGAGAGAAAGCGGAGCGCCAGAG 1197
Dh 1379 GAGCGCTTGGCGCAGCTGAGCGCGCGAGCAGAGAGAGAGAGAGGCTGAGCGCAGAG 1438
Qy 1198 CAGGAGCCCAAGCGGAGCTGAGCTGAGAGAGCACTGGAAGAGAGGGAGCTGAG 1257
Dh 1439 CAGGAGCCCAAGAGCACTGAGAGCTGAGAGCACTGGAAGAGAGGGAGCGGAGCTAGAA 1498
Qy 1258 CGCAGCAGAGAGAGAGAGAGAGAGATGAGAGCGCGAGCGCCCAAAACGGGAA 1317
Dh 1499 CGCAGCAGAGAGAGAGAGAGAGAAATTTGAGAGCGCGAGGCTCCAAAACGGGAA 1558
Qy 1318 CTGGAAG 1377
Dh 1559 CTGGAAG 1618
Qy 1378 AGGAG 1437
Dh 1619 AGGAG 1678

Qy 1438 GAGTTAGAGCTCTGAATGACAAAACATCAGCTAGAGAGAAACTTCAGATATCAGG 1497
Dh 1679 GAATTAGAGCTCTGAATGATTAATAACATCACTAGAGAGGAAACTTCAAGATATCAG 1738
Qy 1498 TGTGAGCTGGCAACCCAGAGCAAGAAATTTGAGAGCAAGCAAAAGTCTAGAGCTAAGA 1557
Dh 1739 TGTGAGTGTGACCAACCAAGCAAGAAATTTGAGAGCAAAACAAATCTAAGAGTTGAGA 1798
Qy 1558 ATTGCTGAAATCACCCACTTAACAGCAGAGTGGAGAGATCTCAGCAAAATGCTTGAAGA 1617
Dh 1799 ATTGCCAAATCACCCACTTAACAGCAACATTAACAGAAATCTCAGCAAAATGCTTGAAGA 1858
Qy 1618 CTATATCCAGAGAAACAGATATCTCAGTACAGTAAACAGTCCAGCAAAACAGTTTG 1677
Dh 1859 CTATATCCAGAGAAACAGATATCTCAGTACCAATTAACAGTTCCAGCAAAACAGTTTG 1918
Qy 1678 CATAGAGCTGGCTTCTTACCTCAAAAGAGCTTGGAAACCAAGAGAGCTGGCCGCGAG 1737
Dh 1919 CACAGAGTTCACTTGTATTAACATTAAGAGCTTGAAGCAAAAGAACTAGCTCGGAG 1978
Qy 1738 CAGCTCCGAGCAGCTGAGCAGAGTGGAGAGAGACCAAGTCAAAAGTGCAGAGATT 1797
Dh 1979 CACCTACAGAGCAACCTGATGAGAGTGGAGAGAGAACTAGATCAAAACCTACAGAGATT 2038
Qy 1798 GATGTTTCAACCAACAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGCACTTCAG 1857
Dh 2039 GATATTTCAATATACAGCTGAGAGAACTAAGAGAAATACAAATTAAGCAAACTCCAG 2098
Qy 1858 AAGCAGAGTCTCCGAGAGCAGCCGCACTGAGCAGAAACAGCAGAGAGAGAGAGCTG 1917
Dh 2099 AAGCAAAAGTCAAGAGAGCTGAACTGAACGACTGAACAGAAACAGAAACCAAGATCATA 2158
Qy 1918 GAGTTAGAGAGCAAAAGAGAGAGCTCAGAGACAGATTCAGGAAAGAGCAAGCAATGG 1977
Dh 2159 GAATTTAGAGAGCAAAAG 2218
Qy 1978 CTGAGCAGTGTGAGCAGAGAG--GAGCAGCAGCGCCCGGAAACCCAGAGAGAGAC 2034
Dh 2219 CTGAGCAGTGTGAGCAG 2278
Qy 2035 AGACTGAAGAGAGAGAGAGAGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
Dh 2279 AAACCTGAAAGAGAGAGAGAGAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2338
Qy 2095 ATGCAAGCAAGCAGAGTGGGCTTTTCATCCGATCAGAGAGCAGTAAAGTGGCCACC 2154
Dh 2339 GCAAGAGCAAGAGTGGGCTGGCTTTTCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2398
Qy 2155 CAGGACCCCTGCTTACCAAG 2214
Dh 2399 CAGGACCCCTGCTTACCTGAG 2458
Qy 2215 AAAGTGTATTTACCGAGAGCTGTACCCCTTTGAATCCAGAAAGTCAAGATGATACC 2274
Dh 2459 AAAGTGTGTATTTACCGAGAGCTGTACCCCTTTGAATCCAGAAAGTGTGAATTAAC 2518
Qy 2275 ATCCAGCAGAGAGATATAGTATGATGTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2334
Dh 2519 ATCCAGCAGAGAGATATAGTATGATGTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2578
Qy 2335 GAGAGAGAGCTGAAAG 2394
Dh 2579 GAGAGAGAGATTAAG 2638
Qy 2395 GAAATAGAGTTCCCACTCCAGAGCAACAGTGAACGATGATGATGATGATGATGATGAT 2454
Dh 2639 GAAATAGAGTTCCCGCTCCAGTGAACAGTGAATGATGATGATGATGATGATGATGATGAT 2698
Qy 2455 AAACCTGCTGCTGAG 2514
Dh 2699 AAACCTGCTGCTGAG 2758

QY 2515 ACCGCCAACAACCTGGGAGAGCTTCACTTCCAGCTGGGCGCAGAGCTCAAGAGAGAGCA 2574
 DB 2759 ACCCTTAATATGATGGGCGGAGCTTCACTTCCAGCTGGGCGCAGAGCTCAAGAGAGCA 2818
 QY 2575 GAAACGACAACTGGGAGAGTACGTGGGCGGCTCAGCCTTCTCTGACCGTACTAGTGTGCG 2634
 DB 2819 GAAACGATTAACCTGGGAGTGTGGGAGCGCCAGCCTCTCTCACTGCTTCAAGTGTGCGG 2878
 QY 2635 CAGTTACGCGAGATCAGCTTCTTACCCAGCAGCAGCAGCTGCTGCTCTCTCTCTCTCC 2694
 DB 2879 CAGTTAAGGAGAGGCTCCGCTTCTTCTCAAGCAGCAGCAGCTGCTGCTCTCTCTCTCT 2938
 QY 2695 GTCTGGGCGAGGAGTGAAGAGGCTCAAGCAGCAGCAGCTGCTGCTCTCTCTCTCTCT 2754
 DB 2939 GTCTAGGCGAGGAGTGAAGAGGCTCAAGCAGCAGCAGCTGCTGCTCTCTCTCTCTCT 2998
 QY 2755 GCCAAGAAAGACACCACTTAATTTTACAAAAGTACGCTCTACCGTTCTGGAACAG 2814
 DB 2999 GCCAAGAAAGACACCACTTAATTTTACAAAAGTACGCTCTACCGTTCTGGAACAG 3058
 QY 2815 CAAAGCATGTGGGTGGTGGAGAGTCAAGGTCAGAGGAGGAGTGTCTCCCAAGCTCTAC 2874
 DB 3059 CAAAGCATGTGGGTGGTGGAGAGTCAAGGTCAGAGGAGGAGTGTCTCCCAAGCTCTAC 3118
 QY 2875 GTGAAATCTATTTCAAGGCGCCGTAAGGAAATCCAAAGCATGCTGAGCTGCTCTGAA 2934
 DB 3119 GTGAAATCTATTTCAAGGCGCCGTAAGGAAATCCAAAGCATGCTGAGCTGCTCTGAG 3178
 QY 2935 AGTCTGCTAGTCTTAAGAGAGTGTCTCCCGCGCGCAGAGCAGCAGTCTCCCGAGAA 2994
 DB 3179 AGTCTGCTAGTCTTAAGAGAGTGTCTCCCGCGCGCAGAGCAGCAGTCTCCCGAGAA 3238
 QY 2995 GAGTTATTCATGTACATACGAGAGTCTGAGCAGAGAGATTTAATCTTTCAAGCA 3054
 DB 3239 GAGTTATTCATGTACATACGAGAGTCTGAGCAGAGAGATTTAATCTTTCAAGCA 3298
 QY 3055 GGGGATGTATGTGTGTACAGAAAGATGTGATCTGTGTGAGCGGAAAGGTGGCGGAC 3114
 DB 3299 GGGGATGTATTTGTGTGTACAGAAAGATGTGATCTGTGTGAGCGGAAAGGTGGCGGAC 3358
 QY 3115 AAGTCCGAGCTCTCCCTTCACTATGTAGGCTTAAAGATTCAGAGGCTCTGAACT 3174
 DB 3359 AAGTCCGAGCTCTCCCTTCACTATGTAGGCTTAAAGATTCAGAGGCTCTGAACT 3418
 QY 3175 GCTGGGAAAAAGGAGGTTTGAAGAAAAAACTGAAATTCGACGTTATTCCTCTAC 3234
 DB 3419 GCTGGGAAAAAGGAGGTTTGAAGAAAAAACTGAAATTCGACGTTATTCCTCTAC 3478
 QY 3235 GCTGCTACTGCTCCGGAACAACCTCAGCTGGCTCTGGGAGCTGATTCGATCCGGA 3294
 DB 3479 ACCGCAACCGGCGCGAGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3538
 QY 3295 AAGAACCAAGTGTGTGTGGAAGAACTGCAAGCTCGAGGAAAAAGCCAGATA 3354
 DB 3539 AAGAACCAAGTGTGTGTGGAAGAACTGCAAGCTCGAGGAAAAAGCCAGATA 3598
 QY 3355 GGGTGTGTCCAGAAATTAATGCAACTTCTAAGCCCGGAAACCAACAAATTCACCCCA 3414
 DB 3599 GGGTGTGTCCAGAAATTAATGCAACTTCTAAGCCCGGAAACCAACAAATTCACCCCA 3658
 QY 3415 ACTGAGTACCAAGACCGCAGCTGACAGCAGCTGCTGAGTGTGAGTGTGAGTGTACAT 3474
 DB 3659 ACAAGGCACTTAAGTCAACAGCATTAAGCGCAGTGTGCGAGTGTGAGTGTACAT 3718
 QY 3475 TACACCGCCAGAAAGATGACGAACTGCTTCAAGCAAGGCGCAGATTCATCAAGTCTC 3534
 DB 3719 TACACCGCCAGAAAGATGACGAACTGCTTCAAGCAAGGCGCAGATTCATCAAGTCTC 3778
 QY 3535 AACAGAGGAGCCCGGAGCTGTGGAAGAGAGAGTGTGAGGAGTGTGAGTGTGAGTGTGAG 3594
 DB 3779 AACAGAGGAGCCCGGAGCTGTGGAAGAGAGAGTGTGAGGAGTGTGAGTGTGAGTGTGAG 3838
 QY 3595 TCCAATTATGTAAAGCTGACCAAGACATGAGCCCGCAGCAATGA 3642

DB 3839 TCCAATTATGTAAAGCTGACCAAGACATGAGCCCAAGCAGCAATGA 3886
 RESULT 6
 AA234571
 ID AA234571 standard; cDNA; 5458 BP.
 XX
 AC AA234571;
 XX
 DT 01-FEB-2000 (first entry)
 DE
 DE Human SH3D1A cDNA clone 21.
 XX
 KW SH3D1A gene; human; Down's syndrome; leukemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lisencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 267..3929
 FT /*tag= a
 PN MO9953062-A2.
 XX
 PD 21-Oct-1999.
 XX
 PF 16-APR-1999; 99MO-US08371.
 XX
 PR 16-APR-1998; 98US-0082007.
 XX
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 PI Korenberg JR, Chen X;
 XX
 DR WP1; 1999-633829/54.
 DR P-PSDB; AAY32155.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukemia -
 PS
 PS Claim 2; Fig 8; 99pp; English.
 XX
 CC This is the nucleotide sequence of full-length cDNA (clone 21)
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
 CC that contributes to the development of platelets and the
 CC pathogenesis of leukemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain
 CC (see AA234570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet
 CC disorder, acute leukemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions,
 CC microcephaly, lisencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.

XX	Sequence	5458 BP;	1671 A;	1201 C;	1307 G;	1279 T;	0 other
SQ							

Query Match	77.7%	Score 2831	DB 20	Length 5458
Best Local Similarity	86.5%	Pred. No. 0		
Matches 3167; Conservative	0	Mismatches 475	Indels 21	Gaps 3

Qy	1	ATGGCTCAGATTTCCCAACACCTTTGCGGTAGGCTGGATGTCTGGGCACTAACTGTGGAG	60
Dp	267	ATGGCTCAGTTTCCAAACCTTTTGGTGGAGGCTGGATATCTGGGCCATTACTGTAGAG	326
Qy	61	GAAGGGCCCAAGCATGACCCGACAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATTACT	120
Dp	327	GAAGAGCCGAGGACATGTATCGACAGTTCCATAGTTTAAAGCAATATCTGGATTCATTACT	386
Qy	121	GGTATCAAGGAGAGAACTTTTTTTCATCTGGGGTACCTCAGGCTGTCTTAGCACA	180
Dp	387	GGTATCAAGCTAGAAACTTTTTTTTTCATCTGGGGTACCTCAGGCTGTCTTAGCACG	446
Qy	241	GCCATGAAGCTTATCAACTGAAGCTTCAAGGATATCAGTCCCCCTCCACACTTCCCCCT	300
Dp	507	GCTATGAACCTTATCAACTGAAGCTTCAAGGATATCAGTACCCTCTCAGTTCCCCCT	566
Qy	301	GTCATGAAGCAGCAACACAGTGGCTATTTCACATGACACACAGACTTTGGTATAGAGGAT	360
Dp	567	GTCATGAAGCAGCAACCAAGTTGCTATTTCACAGGACACAGACTTTGGTATAGGAGTATC	626
Qy	361	GCTATGACATGCCACACTCAGACGTGTGTCTCTGTGCCAATGGGCTCCATTCCAGTTGTT	420
Dp	627	GCCAGCATGCGCACCGGCTTAAAGCTGTGTGCTCCAGTCCCAATGGGATCCATTCCAGTTGT	586
Qy	421	GGAAATGCTCCACACCTTAAATATCTCTGTGCCCTCCAGACAGAGTGGCTCCCGTGGCTAAC	480
Dp	687	GGAAATGCTCCCAACCTTAAGTATCTTCTGTGCCCAAGCAAGCTGTGGCCCCCTGGCTAAC	746
Qy	481	GGGGCTCTCTCCGTCATACAGCCTCTGGCTCTGTGGCATCTTGACGCCACATGGCCA	540
Dp	747	GGGGCTCTCCCTGTATTACAACTCTGGCTGCACTTGTGCTATCTTGACGCCACATGGCCA	806
Qy	541	AAGGCTCTCTCCCTTCCAGCAATCTGGTCCAGGGTCACAATTAACTTAAGTTACAGAG	600
Dp	807	AAGGCTCTCTCTTATGTATATCTGGTCCAGGGTCACAACTTAACTTAAGTTACAGAG	866
Qy	601	GCACAATCATTCATGTGTGCGCAGCGCCCCCTCCAGCAGCAAGATGGGCTGTGCTCAGTCA	660
Dp	867	GCAAGTCATTTTATGTGTGCGCAGTGTGCCACCAAGTGGCAGATGGGCTGTCTCAGTCA	926
Qy	661	TCAAGGCTGAATTCAGGAGCTTATTCACAGCCACGACGACAAACTATAGTGTGAGACATTA	720
Dp	927	TCAAGACTGAATTCAGGAGCTTATTCACATAGTCATGACAAACTATAGTGTGAGACATTA	986
Qy	721	ACAAGTCCCGAGGAGAACTATTTCATGTCATCAAGTTTACCCAGGGCTCAGCGGT	780
Dp	987	ACAAGTCCCGAGGAGAACTATTTCATGTCATCAAGTTTACCAAGGCTCAGCGGT	1046
Qy	781	TCAATATGAATCTTTCTGCACTTTGATTCAGATGAGAAAACTCACTGCAAGAAATTTATC	840
Dp	1047	TCAATATGAATCTTTCTGCACTTTGATTCAGATGAGAAAACTTACAGCAAGAAATTTATC	1106
Qy	841	CTAGACTATGACCTAATTTGATGTGGCAGTGTCTGTGCAGCACTGCGCGGCTCTGCT	900
Dp	1107	CTAGCAATGACCTCATTAATGTAGTGTGTGGCAACCACTGCACTGTCTCTGCT	1166
Qy	901	CCAGAAATACATCCCTCTCTTCTTCCAAAGAGTTCGCTCCGACAGTGGATGTCGCTCAT	960
Dp	1167	CCAGAAATACATCCCTCTTCTTTTAAAGAGTTCGATCTGCGCAGTGTATATCTGTCTAT	1226
Qy	961	AGCTCTTCTTGTGTGATCAGAGGCTGTGAGAGCCGCTGACAGAGATGACGACG	1020

Db	1227	AGCTCCACATCTGTGAGATCAAGAGCGTACCAAGAGAAACAGGTTTATGAAGATGAACAACA	1268
Oy	1021	C---CAGAGAAGAATCTGCTGTGACATTTGAAGATAGAGCGGAGAGATCTTGCAGCGA	1077
Db	1287	CAATTAGAAAAAGAAATTACCTGTATACGTTTAAAGATTAAGAACCGGGAGAACTTTGAAACGT	1346
Oy	1078	GGCAGTGTGGAGCTGGAGAGAGCGCCGCCCAAGCGCTCTTGAGACACAGCGCAAGAGAGAG	1137
Db	1347	GGCAACCTGGAACTGGAGAGAAACGAAGGCAAGCTCTCTTGAAACACAGCGCAAGAGAGAG	1406
Oy	1138	GAGCGGTGTGGCTCAGCTGGAGCGCGCCGAGCAGAGAGAGAGAAAGACCGGGAGCGCAGAG	1197
Db	1407	GAGCGCTGTGGCCAGCTGGAGCGCGCGCGAGACAGAGAGAGAAAGAGACGTGAGCGCCAGAG	1466
Oy	1198	CAGAGGCGCAAGCGCGCAGCTGGAGCTGGAGAAACAGCTTGAGAGACAGCGGAGCTGGAG	1257
Db	1467	CAAGAGCGCAAAAGACAACTGGAATCTGGAGAGCAACTGAAAGACAGCGGGAGCTTAGAA	1526
Oy	1258	CGGCAGCGAGAGAGAGAGAGAGAGAGATGAGAGGCGCGAGCGCCGCAAAACGGGAA	1317
Db	1527	CGGAGAGAGAGGAGGAGAGAGAGAAATTTGAGAGCGAGAGAGCTGCAAAACGGGAA	1586
Oy	1318	CTGGAAGGCGAGCGACAACTTGAATGGGAACGGAAACGGAGACAGAACTCCCTGAATAG	1377
Db	1587	CTTGAAAGGCAACGACAACTTGTGATGGAGAACGGATTCGAAGCGCAAGAACTTAAATCAA	1646
Oy	1378	AGGAACAAGAGAGAGAGAGGAGCAACCTGTGATCTTGAAGGCAAGAGAGAGACTTGGAGTTT	1437
Db	1647	AGGAACAAGAGACAAAGAGGACATATGTTGTATCTGAAGAACAAAGAAAGACTTTGSAATTT	1706
Oy	1438	GAGTTAGAGCTCTGAATGACAAAAAGCATCAGCTTGAAAGAAAATTCAAGATATCAGG	1497
Db	1707	GAATTAGAGCTCTTAATGTATAAAAGCATCAACTGAAGGGGAACTTCAAGATATCAGA	1766
Oy	1498	TGTGACATCGGAACCCAGAGGCGCAAGAAATTTGAGAGCGAAACAAGCTTAGAGAGCTAGA	1557
Db	1767	TGTGATTTGACACCCAAAGGCAAGAAATTTGAGAGACAAACAAATCTTAGAGAGTTGAGA	1826
Oy	1558	ATTGCTGAAATCAACCACTTACAGCAGCAGTTTGACAGATCTCAGCAATGCTTGGAGAA	1617
Db	1827	ATTGCGGAATCAACCACTTACAGCAACAAATTAACAGGAATCTCACCAATGCTTGGAGAA	1886
Oy	1618	CTTATTTCCAGAGAAACAGATATCTCAGTGAACAGTTTAAACAAGTCCAGCAAGACAGTTTG	1677
Db	1887	CTTATTTCCAGAAAAACAGATATCTCAGTGAACCAATTTAAACAAGTTCACAGAAACAGTTTG	1946
Oy	1678	CATGAGACACTCGCTTCTTACCTCAAAAGAGCTTGGAGCAAGAGAGAGAGAGCTGCCCGAG	1737
Db	1947	CACGAGATTTACTTGTTTAACAATTAAAGAGCTTTAGAAAGCAAAAGAACTAGCTCGGAG	2006
Oy	1738	CAGTCTCGGAGAGAGCTGGAACGAGTGGAGAGAGAGAACAGAGTCAAACTGACAGAGATT	1797
Db	2007	CACCTACGAACCAACTGATGAAGTGGAGAAAGAACTAGATCAAACTACAGAGATT	2066
Oy	1798	GATGTTTTCACCAACAGCTGAAGAACTGAGAGAGATCATAGCAAAACAGCAACTCCAG	1857
Db	2067	GATATTTTCAATATACGTGAAAGAACTTAAGAGAAATCACAATATGCAACAACCTCCAG	2126
Oy	1858	AAGCAGAGGTCCTTGGAGGACGGCGCATCTGAACCAAGAAAGACAGAGAGAGAGAGACTTG	1917
Db	2127	AAGCAAAAGTCCATGAGAGGCTGAAACCATGGAACAGAAAGAACGAAGTACTATA	2186
Oy	1918	GAGTTAGAGAGCAAAAGGAAGCGCTCAAGACAGATTCAGGAAGGGGACAAAGCATGG	1977
Db	2187	GAATTTAGAAAAACAAAGAGAGAGCCCAAGACAGAGCTTCAGGAAGGGGACAAAGCATGG	2246
Oy	1978	CTGAGAGCATGTGACAGAGAG--GAGCAGCACGCGCCCGGAAACCCCAAGAGAGAGC	2034
Db	2247	CTGAGAGCATGTGACAGCAGGAGACGAGCATCAAGACCAAGAAATCTCAAGAGAGGAA	2306
Oy	2035	AGACTGAAGGGAGAACAGTGTCAAGAAAGAGAGCGGAGAGAGAGAGCGCAACCGGAA	2094
Db	2307	AAACTGAAAAAGGAGAGAGTGTCAAAAGAAAGAGAGTGGCAGGAAAAAGCGCAACAGAA	2366

QY 2095 ATGCAAGACAGAGAGTGGCTTTTCCATCCGATCAGAGACCACTAAGCTGCCACC 2154
 Db 2367 GCACAAAGACAGAGTGGCTTTTCCATCAACCAAGAACAGCTAAGCCAGCTGTC 2426
 QY 2155 CAGGACCTCTGCTACACAGAGAAAGCCCGCTTACATTTTCCAGAGAGGTGA 2214
 Db 2427 CAGGACCTCTGCTACACAGAGAAAGGTCCATTACATTTTCCAGAGAGAAATGA 2486
 QY 2215 AAAGGTATATTAACGAGAGCTGTACCCCTTTGAATCCAGAGTCAAGATGATCAAC 2274
 Db 2487 AAAGGTATATTAACGAGAGCTGTACCCCTTTGAATCCAGAGTCAAGATGATCAAC 2546
 QY 2275 ATCCAGCCAGAGATATAGTCAAT-----GTTGATGAAGAGCCAGCTGGA 2319
 Db 2547 ATCCAGCCAGAGATATAGTCAAT-----GTTGATGAAGAGCCAGCTGGA 2606
 QY 2320 GAGCCAGATGGCTTGGAGAGAGCTGAAGAGGAGATGGTTCCCTGCAAACTAT 2379
 Db 2607 GAGCCAGATGGCTTGGAGAGAGATTAAGAGAGAGAGAGAGAGGTGTTCCCTGCAAACTAT 2666
 QY 2380 GCAGAAAGATTCAGAGAAATGAGTTTCCACTCCAGCCAAACCACTGACCGATCTGACA 2439
 Db 2667 GCAGAGAAATCCAGAGAAATGAGTTTCCCGCTCCAGTAAACAGTGACTGATCAACA 2726
 QY 2440 TCTGCCCTGCCCCCAAACTGCTCTGCTGAGAACCCCTGCTCTTGGCAGTGAAGCTCT 2499
 Db 2727 TCTGCCCTGCCCCCAAACTGCTCTGCTGAGAACCCCGCTTGGCAGTGAAGCTCT 2786
 QY 2500 TCTGAGCCCTCCAGAAACCCCAACCACTGGGAGAGCTTCAAGTTCAGCTGGCCAGCAGC 2559
 Db 2787 TCAAGAGCCCTCCAGAGCCCTTAATAGTGGGCGAGCTTCAAGCTTCAAGTGGCCAGCAGC 2846
 QY 2560 TCAAGAGAGCCAG 2619
 Db 2847 ACAGATGAG 2906
 QY 2620 GTACCTAGTGGCTGGCCAGATACGAGAGATCAGCCCTTACCCAGCAGCAGCAGCTGCG 2679
 Db 2907 GTTCCAGAGTGGCCAGATACGAGAGAGATCAGCCCTTACCCAGCAGCAGCAGCTGCG 2966
 QY 2680 TCCCTCCCATCTCCCGTCTGGGCGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGG 2739
 Db 2967 TCCCTCCCGTCTCCGTGTAGGCGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGG 3026
 QY 2740 CTGATCTCTGAGAGAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2799
 Db 3027 CTATATCTTGGAGAGCCAAAG 3086
 QY 2800 ACCGTTCTGGAACAGAGAGATGAGTGGTTGGAGAGAGTTCAGAGTCAAGAGGAGTGG 2859
 Db 3087 ACCGTTCTGGAACAGAGAGATGAGTGGTTGGAGAGAGTTCAGAGTCAAGAGGAGTGG 3146
 QY 2860 TTCCCAAGTCTTACGTGAATCTATTTCAAGGCGCGTAAAGAAATCACAAGAGTCTGAT 2919
 Db 3147 TTCCCAAGTCTTACGTGAATCTATTTCAAGGCGCGTAAAGAAATCTAACAAGTCTGAT 3206
 QY 2920 ACTGCGCTTCTGAGAGTGGCTGAGTCTAAAGAGAGTGGTTCCCGGCGCCAGAGCCA 2979
 Db 3207 TCTGTTCTTACAGAGTCTGCTAGTCTAAAGAGAGTGGCTTCCAGAGAGCCAGCCG 3266
 QY 2980 GCCATTCCCGAGAGAGATTTATTCATGTACATACAGAGATTTCTGAGCAAGAGAT 3039
 Db 3267 GTGTTTCCGAGAGAGATTTATTCATGTACATACAGAGATTTCTGAGCAAGAGAT 3326
 QY 3040 TTAACCTTTTACAGAGAGAGAGTGTATTTGTTTACAAAGAAAGTGTGAGTGTGAGC 3099
 Db 3327 TTAACCTTTTACAGAGAGAGAGTGTATTTGTTTACAAAGAAAGTGTGAGTGTGAGC 3386
 QY 3100 GGAACGAGTGGAGAGAGTCCGAGAGTCTTCCCTTCAACTATGTGAGAGCTTAAAGATTTCA 3159
 Db 3387 GGAACGAGTGGAGAGAGTCCGAGAGTCTTCCCTTCAACTATGTGAGAGCTTAAAGATTTCA 3446

QY 3160 GAGGCTCTGGAACCTGCTGGGAAACAGGAGATTAGAGAAAAAACTGAATTTGCCAG 3219
 Db 3447 GAGGCTCTGGAACCTGCTGGGAAACAGGAGATTAGAGAAAAAACTGAATTTGCCAG 3506
 QY 3220 GTTATGCTTCTCAAGCTGTACTGTGCTCCGAACTCACTGCTCTTGGGAGCTTG 3279
 Db 3507 GTTATGCTCTATACACCGCCAGCCCGGAGCAGCTCACTCTCGCTGTGCTGCTG 3566
 QY 3280 ATTGTGATCCGGAAGAAAGAAACCAAGTGTGATGGTGGAGAGAGAACTGCAAGCTCGAGG 3339
 Db 3567 ATTGTGATCCGGAAGAAAGAAACCAAGTGTGATGGTGGAGAGAGAACTGCAAGCTCGAGG 3626
 QY 3340 AAAAGGCGCAGATAGGAGTGGTTTCCAGCAAAATATGTCAAACTTCAAGCCCGGAGACA 3399
 Db 3627 AAAAGGCGCAGATAGGAGTGGTTTCCAGCAAAATATGTCAAACTTCAAGCCCGGAGACA 3686
 QY 3400 AGCAAAATCAACCCCACTGAGCTACCAAGAGCCGAGTGTGAGCAGCAGTGTGCCAGGTG 3459
 Db 3687 AGCAAAATCACTCCACAGAGAGCCCACTTAAGTCAACAGCATTAAGCGCAGTGTGCCAGGTG 3746
 QY 3460 ATGGAGATGTACGATTAACAGCCGCGAGAGATGAGAACTTCAAGCAAGAGCCAG 3519
 Db 3747 ATGGAGATGTACGATTAACAGCCGCGAGAGATGAGAACTTCAAGCAAGAGCCAG 3806
 QY 3520 ATCATCAACGCTCTCAACAGAGAGAGCCGAGTGTGAGAGAGAGAGTCACTGGGCAA 3579
 Db 3807 ATCATCAACGCTCTCAACAGAGAGAGCCGAGTGTGAGAGAGAGAGTCACTGGGCAA 3866
 QY 3580 GTTGGGCTCTTCCCATCTCAATTAATGTAAGCTGACCAACAGAGAGAGAGAGAGAGAGAG 3639
 Db 3867 GTTGGGCTCTTCCCATCTCAATTAATGTAAGCTGACCAACAGAGAGAGAGAGAGAGAGAG 3926
 QY 3640 TGA 3642
 Db 3927 TGA 3929

RESULT 7
 AAS84763
 ID AAS84763 standard; cDNA; 7435 BP.
 AC AAS84763;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #20567.
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN W0200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HISE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG20576.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX

PS Claim 1; SEQ ID No 20567; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 other;

Query Match 68.4%; Score 2491.6; DB 23; Length 7435;
Best Local Similarity 84.2%; Pred. No. 0;
Matches 3112; Conservative 0; Mismatches 529; Indels 53; Gaps 25;

QY 1 ATGGCTCAAGTTTCCACACCTTTCGGTGTAGCTGTGATGTCGGGCCATACTGTGAG 60
DB 261 ATGGCTCAAGTTTCCACACCTTTCGGTGTAGCTGTGATGTCGGGCCATACTGTGAG 322
QY 61 GAAAGGCCAACAGTACGAGATGTTCTTACCTGAGCCATGCGGGATTTATTACT 120
DB 323 GAAAGGCCAACAGTACGAGATGTTCTTACCTGAGCCATGCGGGATTTATTACT 382
QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTCCTCAGCTGTCTTACACAA 180
DB 383 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTCCTCAGCTGTCTTACACAA 442
QY 181 ATATGGGCGCTAGCGGACATGAATAAGATGAGATGATCAAGTATTTTCATA 240
DB 443 ATATGGGCGCTAGCGGACATGAATAAGATGAGATGATCAAGTATTTTCATA 502
QY 241 GGCATGAAGCTTATCAAACTGAAGCTACAGATATCAGCTCCCTCCACACTTCCCT 300
DB 503 GGTATGAACTTATCAAACTGAAGCTACAGATATCAGCTCCCTCCACACTTCCCT 562
QY 301 GTCATGAACAGCAACAGTGTGCTATTTTCAGTGACAGCATTTGGTATAGAGAGATT 360
DB 563 GTCATGAACAGCAACAGTGTGCTATTTTCAGTGACAGCATTTGGTATAGAGAGATT 622
QY 361 GGTAGCATGCACTCACTCAGCTGTGCTCTGTGCAATGGAGCTTCATT-CCAGTTGT 419
DB 623 GCGAGATGCCACCGCTTACAGCTGTGCTCCAGTGCCAAATGGATTCATCCAGTTGT 682
QY 420 TGGAAATGTCACACCTTAGTATCTTCTCTCCCTCCAGAGAGAGTCTCCCTGGCTAA 479
DB 683 TGGAAATGTCACACCTTAGTATCTTCTCTCCCTCCAGAGAGAGTCTCCCTGGCTAA 742
QY 480 CGGGGCTCTCCGTCATACAGCTCTGCTGCTGTTGGCATCTCCGAGCCACATGGCC 539
DB 743 CGGGGCTCTCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802
QY 540 AAGAGTTTCTCTTACAGCAGATCTGTGTCAGAGTCACAATTTAAACATTAATTACAGAA 599
DB 803 AAGAGTTTCTCTTACAGCAGATCTGTGTCAGAGTCACAATTTAAACATTAATTACAGAA 862
QY 600 GGCACATATTCATGATGTGCGGAGCGCCCTCC-AGCAGAGAAATGGGTGTGCTCAGT 658
DB 863 GGCACATATTCATGATGTGCGGAGCGCGCTCCACCAAGTGGCAGAGTGGGTGTCTCAGT 922

QY 659 CATCAAGCTGAATAACA---GGCAGTTATTCAACGCCAGCAAAAC--TATGATGG 713
DB 923 CATCAAGCTGAATAACA---GGCAGTTATTCAACGCCAGCAAAAC--TATGATGG 982
QY 714 ACATTT--AACAGTCCCCCGAGCAAGACTATTCTCATGCAATCAAGTTTACCAGGCT 771
DB 983 ACATTT--AACAGTCCCCCGAGCAAGACTATTCTCATGCAATCAAGTTTACCAGGCT 1042
QY 772 CAGTGGCTTCAATATGAAATCTTTTGCATTTGATCAATGAAATGAAATCACTGAGAA 831
DB 1043 CAGTGGCTTCAATATGAAATCTTTTGCATTTGATCAATGAAATGAAATCACTGAGAA 1102
QY 832 GAATTTATCTAGCTATGCACTTAATGATGTTGATGTCATGTCGTCAGCCATCCGCC 891
DB 1103 GAATTTATCTAGCTATGCACTTAATGATGTTGATGTCATGTCGTCAGCCATCCGCC 1162
QY 892 GTCTGCTCTCCAGATATCATCTCTCTTCTTCAAGAGTTGCTTCCGCACTGGGATG 951
DB 1163 GTCTGCTCTCCAGATATCATCTCTCTTCTTCAAGAGTTGCTTCCGCACTGGGATG 1222
QY 952 TCGGTCAATAGCTCTTCTTGTGATCAGAGGCTGCTGAGAGCCGTGTCAGAGAT 1011
DB 1223 TCGGTCAATAGCTCTTCTTGTGATCAGAGGCTGCTGAGAGCCGTGTCAGAGAT 1282
QY 1012 GAGCAGCAGC---CAGAGAAATCTGCTGTGACATTTGAAATGAAGCCGAGAAAC 1068
DB 1283 GAGCAGCAGC---CAGAGAAATCTGCTGTGACATTTGAAATGAAGCCGAGAAAC 1342
QY 1069 TTGAGCAGAGCAGTGTGAGCTGTGAGAACGCCGCCAGCGCTTTTGGAGCAGCGC 1128
DB 1343 TTGAGCAGAGCAGTGTGAGCTGTGAGAACGCCGCCAGCGCTTTTGGAGCAGCGC 1402
QY 1129 AAGAGCAGAGCGGTTGGCTGCTGAGCGCGCGAGCAGAGAGAGAAAGCGCGAG 1188
DB 1403 AAGAGCAGAGCGGTTGGCTGCTGAGCGCGCGAGCAGAGAGAGAAAGCGCGAG 1462
QY 1189 CGCAGAGCAGAGCGCGCGAGCAGCGCGAGCAGCGCTGAGAGAGCAGCGCAGCG 1248
DB 1463 CGCAGAGCAGAGCGCGCGAGCAGCGCGAGCAGCGCTGAGAGAGCAGCGCAGCG 1522
QY 1249 GAGCTGAGCGGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308
DB 1523 GAGCTGAGCGGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
QY 1309 AAGAGGAACT--GAAAGGAGAGCACTTGAATGGGAACGGAACCGAGACAGAACT 1367
DB 1583 AAGAGGAACT--GAAAGGAGAGCACTTGAATGGGAAGTGGAGAGGCGATCGAGACT 1642
QY 1368 CTTGAATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1427
DB 1643 ACTTAATCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1702
QY 1428 TTTGGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGA 1487
DB 1703 TTTGGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGA 1762
QY 1488 GATATCAGAGTGTGCACTGCGCAACCAAGAGCAAGAAATTTGAGAGAGCAAGACTAG 1547
DB 1763 GATATCAGAGTGTGCACTGCGCAACCAAGAGCAAGAAATTTGAGAGAGCAAGACTAG 1822
QY 1548 AGAGCTAAGAAATTTGCTG--AAATCACCACCTTACAGAGAGAGT--GAGAGAAATTCAG 1603
DB 1823 AGAGCTAAGAAATTTGCTG--AAATCACCACCTTACAGAGAGAGT--GAGAGAAATTCAG 1882
QY 1604 AAA--TGCTTGAAGAGCTTATTTCAAGAGAAAGAGATCTCACTGACAGTTTAAACAG 1660
DB 1883 AAA--TGCTTGAAGAGCTTATTTCAAGAGAAAGAGATCTCACTGACAGTTTAAACAG 1942
QY 1661 TCGAGCAGAAACAGTTTGCATAGAGACTGCTTCTTACCTCAAAAG--AGCTTTGAGAGAG 1719
DB 1943 TCGAGCAGAAACAGTTTGCATAGAGACTGCTTCTTACCTCAAAAG--AGCTTTGAGAGAG 2002

QY 1720 AAGAGCTGGCCCCGAGCAGCTCCGGAGCAGCTGAGCGAGGTGAGAGAGACCAG 1779
 Db 2003 AAGAGCTGGCCCCGAGCAGCTCCAGAGACCAACTGGATGAGTGAAGAAAGAACTAGA 2062
 QY 1780 TCAGAGCTGAGAGATTTGTTTTCACACACAGCTGAGAGAACTGAGAGATACAT 1839
 Db 2063 TCAGAGCTGAGAGATTTGTTTTCACACACAGCTGAGAGAACTGAGAGATACAT 2122
 QY 1840 AGAGAGCAGCACTCCAGAGCAGAGGTCCCTGAGGCGAGCGAGCTGAGAGAGAGAG 1899
 Db 2123 AATAGAGCACTCCAGAGCAGAGAAAGTCCATGAGGCTGAGAGCTGAGAGAGAGAG 2182
 QY 1900 CAGAGAGAGAGAGCTGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1959
 Db 2183 CAGAGAGAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2242
 QY 1960 GAG 2016
 Db 2243 GAG 2302
 QY 2017 AAG 2075
 Db 2303 AAG 2362
 QY 2076 AG 2135
 Db 2363 GAG 2422
 QY 2136 GAG 2195
 Db 2423 AAG 2482
 QY 2196 TTTGAG 2255
 Db 2483 TTTGAG 2542
 QY 2256 AAG 2300
 Db 2543 AAG 2602
 QY 2301 GAG 2360
 Db 2603 GAG 2662
 QY 2361 AAG 2420
 Db 2663 GAG 2722
 QY 2421 AAG 2480
 Db 2723 AAG 2782
 QY 2481 TTTGAG 2540
 Db 2783 TTTGAG 2842
 QY 2541 TTTGAG 2600
 Db 2843 TTTGAG 2902
 QY 2601 GAG 2659
 Db 2903 GAG 2962
 QY 2660 CAG 2719
 Db 2963 CAG 3022
 QY 2720 AAG 2778
 Db 3023 AAG 3082
 QY 2779 TTTAAG 2838

Db 3083 TTTAAG 3142
 QY 2839 GTTCAAG 2898
 Db 3143 GTTCAAG 3202
 QY 2899 AG 2957
 Db 3203 AG 3262
 QY 2958 GAG 3016
 Db 3263 AG 3322
 QY 3017 AG 3072
 Db 3323 AG 3382
 QY 3073 AG 3131
 Db 3383 AG 3442
 QY 3132 TTTCAAG 3191
 Db 3443 TTTCAAG 3502
 QY 3192 TTTCAAG 3250
 Db 3503 TTTCAAG 3562
 QY 3251 AAG 3309
 Db 3563 AAG 3622
 QY 3310 TTTGAG 3369
 Db 3623 TTTGAG 3682
 QY 3370 AATTAAG 3429
 Db 3683 AATTAAG 3742
 QY 3430 ACCGAG 3489
 Db 3743 ACCGAG 3802
 QY 3490 GATGAG 3548
 Db 3803 GATGAG 3862
 QY 3549 GAG 3608
 Db 3863 GAG 3922
 QY 3609 GCTGAG 3641
 Db 3923 GCTGAG 3956

RESULT 8
 AA234570 standard; cDNA; 5199 BP.

AA234570;

01-FEB-2000 (first entry)

Human SH3D1A cDNA.

SH3D1A gene; human; Down's syndrome; leukemia; cancer;
 megakaryocytic abnormality; myeloproliferative disorder;
 platelet disorder; neutral disorder; thrombocytopenia;

DB 3415 ACAGACCCACCTTAAGTCAACAGCATTAGCGGCGAGTGTGCGAGTGTGGAGTACGAC 3474
QY 3475 TACACCGCCAGACGATGACGAACTAGCCTTACGAAAGGCGAGATCATCAACGCTCC 3534
DB 3475 TACACCGCCAGATGATGATGAGTGGCTTCAACAGGCGGATCATCAACGCTCC 3534
QY 3535 AACAGAGAGACCCGAGCTGTGTGAAGAAGAGTCAAGTGGCAAGTTGGCTTTCCCA 3594
DB 3535 AACAGAGAGACCCGAGCTGTGTGAAGAAGAGTCAAGTGGCAAGTTGGCTTTCCCA 3594
QY 3595 TCCAAATTATGTAAAGTGCACACAGCATGACCCGACGACGAAATGA 3642
DB 3595 TCCAAATTATGTAAAGTGCACACAGCATGACCCGACGACGAAATGA 3642
RESULT 9
ID ABK43498 standard; cDNA; 3319 BP.
XX ABK43498;
AC ABK43498;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #78.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN MO20015318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235464.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236338.
PR 29-SEP-2000; 2000US-0236339.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX MPI, 2001-581633/65.

DR P-PSDB; AAB87168.

XX New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 PT food additives or preservatives -

XX Claim 1; SEQ ID No 88; 837pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein. (I) and polypeptides (III) encoded
 CC by (I), are used to treat a medical conditions and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infarction. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 56.1%; Score 2043.8; DB 23; Length 3319;
 Best Local Similarity 85.2%; Pred. No. 0;
 Matches 2324; Conservative 0; Mismatches 387; Indels 18; Gaps 3;

QY 1 ATGGCTAGTTCCACACCTTTCGGGTGAGCTGATCTGGGCACTAAGTGGAG 60
 DB 159 ATGGCTAGTTCCACACCTTTCGGGTGAGCTGATCTGGGCACTAAGTGGAG 218
 QY 61 GAAAGGCCAAGCATGACACGACCTTCTTACCTGAGCCGATAGCGGATTTACT 120
 DB 219 GAAAGGCCAAGCATGACACGACCTTCTTACCTGAGCCGATAGCGGATTTACT 278
 QY 121 GGTATCAAGGAGGAACTTTTTCCTCAATCTGGGTACTCAGCTCTTACGACAA 180
 DB 279 GGTATCAAGGAGGAACTTTTTCCTCAATCTGGGTACTCAGCTCTTACGACAA 338
 QY 181 ATATGGGCTAGCGGACATGAAATACGATGAAAGATGAGTCAAGTGAATTTCCATA 240
 DB 339 ATATGGGCTAGCGGACATGAAATACGATGAAAGATGAGTCAAGTGAATTTCCATA 398
 QY 241 GCCATGAAGCTTATCAAACTGAGCTCAAGGATATAGCTCCCTCCACACTTCCCCCT 300
 DB 399 GCTATGAAACTTATCAAACTGAGCTCAAGGATATAGCTCCCTCCACACTTCCCCCT 458
 QY 301 GTCATGAAACAGCAACGAGTGGCTATTCCAGTGCACACGACATTTGGTATAGAGGAGTT 360
 DB 459 GTCATGAAACAGCAACGAGTGGCTATTCCAGTGCACACGACATTTGGTATAGAGGAGTT 518
 QY 361 GCTAGCATGCCACCACTGACAGCTGTTGCTCTGTCGCAATGGGCTCATTCCAGTTGTT 420
 DB 519 GCCAGCATGCCACCGCTTACAGCTGTTGCTCTGTCGCAATGGGATCATTCCAGTTGTT 578
 QY 421 GGAATGTCACACCTTAGTATCTTCTGTCCTCCAGCAGAGAGCTCCCTGGGCTAAC 480
 DB 579 GGAATGTCACACCTTAGTATCTTCTGTCCTCCAGCAGAGAGCTCCCTGGGCTAAC 638
 QY 481 GGGGCTCTCCGTCATACAGCTCTGCTGCTGGTGGCATCTCTGACGCAACATGGGCCA 540
 DB 639 GGGGCTCTCCGTCATACAGCTCTGCTGCTGGTGGCATCTCTGACGCAACATGGGCCA 698
 QY 541 AAGAGTTCTTCTTCCAGCAGATCTGTCGCGGCACAATTAACACTAGTTACAGAG 600
 DB 699 AAGAGTTCTTCTTCCAGCAGATCTGTCGCGGCACACTAATTAACACTAGTTACAGAG 758
 QY 601 GCACATCATTCGATGTCGCGGCGCCCTCCAGCAGACAGATGGGCTGTCCTAGTCA 660
 DB 759 GCACATCATTCGATGTCGCGGCGCCCTCCAGCAGACAGATGGGCTGTCCTAGTCA 818
 QY 661 TCAGAGCTGAATACAGGAGCTTATTAACAGCCACAGCAAACTATAGTGAACACTTA 720
 DB 819 TCAGAGCTGAATACAGGAGCTTATTAACAGCCACAGCAAACTATAGTGAACACTTA 878
 QY 721 AC-----AGGTCCTCCAGCAAGAACTTCTCATGTCATCAATCAAGTTTACCCAG 768
 DB 879 ACAGGTTCTTATAGTCCCAAGCAAGAACTTCTTATGACGTCAAGTTTACCCAG 928
 QY 769 GCTCAGCTGAGCTTCAATATGGAATCTTTTGAACATTATCAATCAAGTGAATCACTGCA 828
 DB 929 GCTCAGCTGAGCTTCAATATGGAATCTTTTGAACATTATCAATCAAGTGAATCACTGCA 998
 QY 829 GAAAGATTTATCCATGATGACCTAATGATGATGATGATGATGATGATGATGATGATGAT 888
 DB 999 GAAAGATTTATCCATGATGACCTAATGATGATGATGATGATGATGATGATGATGATGAT 1058
 QY 889 CCGGCTCTGCTCCAGATATACATCTCTCTTCAAGAGAGTTGCTCCGGCAGTGGG 948
 DB 1059 CCGGCTCTGCTCCAGATATACATCTCTCTTCAAGAGAGTTGCTCCGGCAGTGGG 1118
 QY 949 ATGTCGCTATTAAGTCTTTCTTGTGTGATGAGGCTGCTGAGAGCCGCTGTCAG 1008
 DB 1119 ATATGTCATTAAGTCAACATCTGTGATGAGGCTGCTGAGAGCCGCTGTTAGAA 1178
 QY 1009 GATGAGCAGCAGC---CAGAGAGAAACTGCTGTGACATTTGAAGATTAAGAGCGGAG 1065


```

Db 1179 GATGAAACAACAATTTAAGAAATTAACCTGTAACCTTTGAAGATTAAGAGGGAG 1238
Oy 1066 AACTTGAGCGAGCGAGCTGTGGAGTGTGAGAACCGCGCAAGCCCTCTTGAGCAGAG 1125
Db 1239 AACTTTGAACGTGGCAACTGTGAGAAACGAGGCAAGCTCTCTGGAACAGCAG 1298
Oy 1126 CGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGG 1185
Db 1299 CCAAGAGACAGAGAGCGCTGGCCCGAGCTGAGCGGGCGAGCAGAGAGAGAGAGCGT 1358
Oy 1186 GAGCGCCAGAGACAGAGAGCGGCGAGCTGTGAGTGTGAGAAAGCAGCTGAGAGAG 1245
Db 1359 GAGCGCCAGAGACAGAGAGCGGCGAGCTGTGAGTGTGAGAAAGCAGCTGAGAGAG 1418
Oy 1246 CGGAGCTGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
Db 1419 CGGAGCTGAGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1478
Oy 1306 GCAAAAACGGGAACTGGAAAAGCAGCGCAACTTGAATGGGAAACGAAACCGAGAGAG 1365
Db 1479 GCAAAAACGGGAACTTGAAGGCAACGACAACTTGAAGTGGGAAACGAGATCGAAGCA 1538
Oy 1366 CTCCTGAATCAAGAGAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
Db 1539 CTACTAAATCAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1598
Oy 1426 ACTCTGAGATTTGAGTTAGAGCTGTAAGCTGTAAGCAAAAGCATCAGCTGAGAGAG 1485
Db 1599 ACTTTGAAATTTGAATTTAAGCTCTAAATGATTAAGATCACTAACTGAAGAGAGAG 1658
Oy 1486 CAGATATCAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1545
Db 1659 CAAAGATATCAGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1718
Oy 1546 AAGAGCTAAGAAATGTGCTGAATCACCCACTTACAGAGAGAGTGTGAGAGATCTCAG 1605
Db 1719 AAGAGCTAAGAAATGTGCTGAATCACCCACTTACAGAGAGAGTGTGAGAGATCTCAG 1778
Oy 1606 ATGCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1665
Db 1779 ATGCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1838
Oy 1666 CAGAACAGTTTGCATAGAGAGAGTGTGCTTCTTACCTCAAAAGAGCTTGTGAGAGAG 1725
Db 1839 CAGAACAGTTTGCATAGAGAGAGTGTGCTTCTTACCTTAAAGAGAGCTTGTGAGAGAG 1898
Oy 1726 CTGGCCCGGAGAGAGTGTGCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAG 1785
Db 1899 CTAGCTCGGAGAGAGTGTGCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAG 1958
Oy 1786 CTGAGAGAGATTTGAGTGTTCACAAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1845
Db 1959 CTACAGAGATTTGATTTTCAATTAATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2018
Oy 1846 CAGCAACTCAGAGAGAGAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1905
Db 2019 CAGCAACTCAGAGAGAGAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2078
Oy 1906 AGGAGAGAGCTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
Db 2079 CGAAAGATCATGAAATTTAGAAAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2138
Oy 1966 GACAGAGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
Db 2139 GACAGAGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
Oy 2023 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
Db 2199 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2258
Oy 2083 GCCAAGCCGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2142

```

```

Db 2259 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2318
Oy 2143 AAGCTGGCCACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2202
Db 2319 AAGCAGAGCTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2378
Oy 2203 CAGAGAGAGTAAAGAGTGTATTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2262
Db 2379 CAGAGAGAGTAAAGAGTGTATTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2438
Oy 2263 GATGATACACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2322
Db 2439 GATGATACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2498
Oy 2233 CCAGATAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2382
Db 2499 CCGGCTGTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2558
Oy 2283 GAAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442
Db 2559 GAGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2618
Oy 2443 GCGCCGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
Db 2619 GCGCCGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2678
Oy 2503 GAGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
Db 2679 GAGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2738
Oy 2563 AACGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
Db 2739 AATGAGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2798
Oy 2623 CCTAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2682
Db 2799 CCAAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2858
Oy 2683 TCCCATCTCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
Db 2859 TCCCATCTCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2887

RESULT 10
AA163825
ID AA163825 standard; cDNA; 3466 BP.
XX
AC AA163825;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 33.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; viruticid;
KW fungicidic; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; anti-allergic; hepatotropic; antidiabetic;
KW anti-inflammatory; anti-ulcer; anticonvulsant; antibacterial;
KW antiparasitic; cardiatic; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
XX
OS Homo sapiens.
XX
PN MO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.

```


CC the encoded proteins (AAM43497-AAM4360) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3466 BP, 1056 A; 829 C; 861 G; 710 T; 10 other;

Query Match 56.1%; Score 2043.8; DB 22; Length 3466;

Best Local Similarity 85.2%; Pred. No. 0;

Matches 2224; Conservative 0; Mismatches 387; Indels 18; Gaps 3;

Qy 1 ATGGCTCAAGTTCACACCTTTGGTGTAGCTGTGATGCTGGGCCATTAAGCTGAG 60
 Db 277 ATGGCTCAAGTTCACACCTTTGGTGTAGCTGTGATGCTGGGCCATTAAGCTGAG 336
 Qy 61 GAAAGGGCCAAACATGACGAGGATTCCTTACCTGAAAGCCGATGCGGAGTTTACT 120
 Db 337 GAAAGGGCCAAACATGACGAGGATTCCTTACCTGAAAGCCGATGCGGAGTTTACT 396
 Qy 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTACTCAGCTGTCTTACAGAA 180
 Db 397 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTACTCAGCTGTCTTACAGAG 456
 Qy 181 ATATGGGCGCTTACCGGACATGAATACATGAGAGATGATCAAGTGAATTTTCCATA 240
 Db 457 ATATGGGCGCTTACCGGACATGAATACATGAGAGATGATCAAGTGAATTTTCCATA 516
 Qy 241 GCCATGAACCTTATCAAACTGAAAGCTACAGATATCACTCCCTCCACACTTCCCT 300
 Db 517 GCCATGAACCTTATCAAACTGAAAGCTACAGATATCACTCCCTCCACACTTCCCT 576
 Qy 301 GTCATGAACACAGCAACCACTGGCTATTTCCAGTGCACACGATTTTGTATAGAGGAT 360
 Db 577 GTCATGAACACAGCAACCACTGGCTATTTCCAGTGCACACGATTTTGTATAGAGGAT 636
 Qy 361 GCTAGCATGCAACCACTGACCTGTGCTCTGTGCCAATGGGCTCCATTCAGTTGTT 420
 Db 637 GCTAGCATGCAACCACTGACCTGTGCTCTGTGCCAATGGGCTCCATTCAGTTGTT 696
 Qy 421 GGAATGTCTCCACCTTATGATCTTCTGCTCCAGAGCAGTGTGCTTCCCTGGCTAAC 480
 Db 697 GGAATGTCTCCACCTTATGATCTTCTGCTCCAGAGCAGTGTGCTTCCCTGGCTAAC 756
 Qy 481 GGGGCTCTCCCTGCTCAAGCCTGCTGCTGCTTGGCCATCTGCAAGCAGCATGAGCA 540
 Db 757 GGGGCTCTCCCTGCTCAAGCCTGCTGCTGCTTGGCCATCTGCAAGCAGCATGAGCA 816
 Qy 541 AAGAGTTCTTCTTACAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 Db 817 AAGAGTTCTTCTTACAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
 Qy 601 GCACATCTATTTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 Db 877 GCACATCTATTTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
 Qy 661 TCAGAGCTGAATATACAGAGATTTTCAACAGCAGCAAGAACTATGATGAGACATTA 720
 Db 937 TCAGAGCTGAATATACAGAGATTTTCAACAGCAGCAAGAACTATGATGAGACATTA 996
 Qy 721 AC-----AGGTCCCAAGGCAAGAACTATTTCTCATGCAATCAAGTTTACCAG 768

Db 997 ACAGGTTCTGTTAGGTCCCAAGAGAACTATTTATGATGATCAAGTTTACCAG 1056
 Qy 769 GCTCAGCTGCTTCAATATGAAATCTTTTCTGACATGATCAAGATGAAATCACTGCA 828
 Db 1057 GCTCAGCTGCTTCAATATGAAATCTTTTCTGACATGATCAAGATGAAATCACTGCA 1116
 Qy 829 GAAGAAATTTATCTAGCTATGACCTTAATGATGTTGCTGCTGCTGCTGCTGCTGCT 888
 Db 1117 GAGAAATTTATCTAGCTATGACCTTAATGATGTTGCTGCTGCTGCTGCTGCTGCTGCT 1176
 Qy 889 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
 Db 1177 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1236
 Qy 949 ATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008
 Db 1237 ATATCTGCTATTAAGCTCAATCTGATGATCAGAGCTACCAAGAGAACCAAGTTTGA 1296
 Qy 1009 GATGAGCAGCAGC--CAGAGAAAGAACTGCTGTGACATTTGAATGAAGAGCGGAG 1065
 Db 1297 GATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1356
 Qy 1066 AACTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1125
 Db 1357 AACTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1416
 Qy 1126 CGCAAGAGCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185
 Db 1417 CGCAAGAGCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1476
 Qy 1186 GAGCGCAGAGCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1245
 Db 1477 GAGCGCAGAGCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1536
 Qy 1246 CGGAGCTGAGCGGCGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 1305
 Db 1537 CGGAGCTGAGCGGCGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 1596
 Qy 1306 GCAAAAAGGAACTGAAAGAGCAGCAGCACTTGAATGGGAGCGGAGCAGAGAGAG 1365
 Db 1597 GCAAAAAGGAACTGAAAGAGCAGCAGCACTTGAATGGGAGCGGAGCAGAGAGAG 1656
 Qy 1366 CTCCTGAATCAGAGGAAAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGAG 1425
 Db 1657 CTCCTGAATCAGAGGAAAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGAG 1716
 Qy 1426 ACTGAGATTTGATGAGTGAAGCTTGAATGACAAAAGATCAAGTGAAGAGAACTT 1485
 Db 1717 ACTTGAATTTGATGAGTGAAGCTTGAATGATGAAGATCAAGTGAAGAGAACTT 1776
 Qy 1486 CAGGATATCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1545
 Db 1777 CAGGATATCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1836
 Qy 1546 AGAGAGTGAAGATTTGCTGAATACCCACTTAAGCAGCAGAGTTCAGAGATTCAGCA 1605
 Db 1837 AGAGAGTGAAGATTTGCTGAATACCCACTTAAGCAGCAGAGTTCAGAGATTCAGCA 1896
 Qy 1606 ATGCTTGAAGATTTTCCAGAGAAACATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1665
 Db 1897 ATGCTTGAAGATTTTCCAGAGAAACATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1956
 Qy 1666 CAGAGAGTTTGCATGAGAGCTGCTTCTTACCTCAAAAGAGCCTTGAAGAGAGAG 1725
 Db 1957 CAGAGAGTTTGCATGAGAGCTGCTTCTTACCTCAAAAGAGCCTTGAAGAGAGAG 2016
 Qy 1726 CTGCGCCGAGCAGCTCCGAGCAGCTGAGCAGAGTGAAGAGAGAGCAGCTCAAG 1785
 Db 2017 CTGAGCTCGGAGCAGCTCAAGAGCAGCTGAGTGAAGTGAAGAGAGAGCAGCTCAAG 2076
 Qy 1786 CTGAGAGATTTGATGTTTTCATCAACCAAGCTGAAGAGAGTGAAGAGATATAGCAAA 1845

```

Db 2077 CTACAGAGATTGATTTTCATATACAGCTGAGGAACTAAGAGAAATACCAATTAAG 2136
QY 1846 CAGCAATCCAGAAAGCAGAGAGTCCCTGAGAGCAGCGGACTGAAGAGAAAGAGCAGAG 1905
Db 2137 CAACAACTCCAGAAAGCAAAAGTCCATGAGGCTGAAGACTGAACAAGAAAGCAAGAA 2196
QY 1906 AGGAAGAGCTGGAGTTAGAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
Db 2197 CGAAGATCATAGATTAGAAAAAAGAAAAAGAAAGAAAGAGAGAGAGAGAGAGAG 2256
QY 1966 GACAAGCAATGGCTGAGAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
Db 2257 GACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2316
QY 2023 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
Db 2317 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2376
QY 2083 GCCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2142
Db 2377 GCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2436
QY 2143 AAGCTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2202
Db 2437 AAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496
QY 2203 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2262
Db 2497 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2556
QY 2263 GATGAGATCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2322
Db 2557 GATGAGATCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2616
QY 2323 CCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2382
Db 2617 CCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2676
QY 2383 GAAAAATGTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442
Db 2677 GAGAAATGTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2736
QY 2443 GCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
Db 2737 GCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2796
QY 2503 GAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
Db 2797 GAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2856
QY 2563 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
Db 2857 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2916
QY 2823 CCAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2882
Db 2917 CCAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2976
QY 2683 TCCCATCTCCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
Db 2977 TCCCATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3005

```

```

KW SH3D1A gene; human; Down's syndrome; leukemia; cancer;
KW megakaryocytic abnormality; myeloproliferative disorder;
KW platelet disorder; neural disorder; thrombocytopenia;
KW haematopoietic disorder; cognitive dysfunction; microcephaly;
KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
KW ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 136..2079
FT FT /*tag= a
FN W09953062-A2.
PD 21-OCT-1999.
PR 16-Apr-1999; 99WO-US08371.
PR 16-Apr-1998; 98US-0082007.
PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
PI Korenberg JR, Chen X;
PI WPI; 1999-633829/54.
DR P-PSDB; AAY32157.
PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukemia -
PS Claim 2; Fig 12; 99PP; English.
XX
XX This is the nucleotide sequence of a cDNA clone, termed clone 5,
XX corresponding to a novel human SH3 gene, termed the SH3D1A gene,
XX that contributes to the development of platelets and the
XX pathogenesis of leukemias, both in general and in particular those
XX involving the megakaryocytic lineage. The SH3D1A gene maps to the
XX small candidate region for low platelets on chromosome 21.
XX Sequencing of 5 different sizes of cDNA clone from foetal brain
XX (see AA234570-74) suggests that at least 3 isoforms exist. The
XX invention provides methods for the diagnosis and treatment of
XX megakaryocytic abnormality, myeloproliferative disorder, platelet
XX disorder, acute leukemia, neural disorders, thrombocytopenia,
XX platelet disorder on chromosome 21, low platelets in deletion for
XX 21, association of gains in chromosome 21 with leukemias, neural
XX abnormalities, dysfunctions and disorders including brain
XX malformations and corresponding cognitive dysfunctions,
XX microcephaly, lissencephaly, and colpocephaly. Methods are also
XX provided for: suppressing cells unable to regulate themselves;
XX screening for a somatic alteration in the SH3D1A gene; monitoring
XX the progress and adequacy of a treatment; monitoring tumour risk
XX progress or megakaryocytic abnormality; myeloproliferative disorder,
XX haematopoietic disorder, platelet disorder or leukemia; and
XX treatment of a subject (including a prenatal subject) having
XX megakaryocytic abnormality, myeloproliferative disorder,
XX platelet disorder, leukemia or neural disorder using a
XX nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX
SQ Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 other;

```

Query Match 41.0%; Score 1495; DB 20; Length 2079;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```

QY 1 ATGGCTAGTTTCCACACCTTTGGTGTAGCTGAGATGCTGGCCATTAAGTGTGAG 60
Db 136 ATGGCTAGTTTCCACACCTTTGGTGTAGCTGAGATGCTGGCCATTAAGTGTGAG 195
QY 61 GAAAGGCGCAAGCATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 196 GAAAGGCGCAAGCATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255

```

QY 121 GGTGATCAAGCGAAGACTTTTTCATCTGGCTTACCTGAGCTCTTACGACA 180
 DB 256 GGGATCAAGCTAGAAACTTTTTCATCTGGCTTACCTGAGCTCTTACGACA 315
 QY 181 ATATGGCGCTGCGGACATGAAATGAAGATGAGATCAAGTGAATTTTCCATA 240
 DB 316 ATATGGCGCTGCGGACATGAAATGAAGATGAGATCAAGTGAATTTTCCATA 375
 QY 241 GCCATGAGCTTATCAAACTGAAAGTACAAAGATATCAGTCCCTCCACATTTCCCT 300
 DB 376 GGTATGAACCTATCAAACTGAAAGTACAAAGATATCAGTCCCTCCACATTTCCCT 435
 QY 301 GTTCATGAACAGCAACCAAGTGTATTTCCAGTGAACAGCATTTGGTATGAGAGGAT 360
 DB 436 GTTCATGAACAGCAACCAAGTGTATTTCCAGTGAACAGCATTTGGTATGAGAGGAT 495
 QY 361 GGTAGATGCGCAACCTGACAGCTGTGCTCTGTCGCAATGGGCTCCATTCAGATTT 420
 DB 496 GCGAGATGCGCAACCTGACAGCTGTGCTCTGTCGCAATGGGCTCCATTCAGATTT 555
 QY 421 GGAATGTCTCAACCTTATATCTTCTGCTCCAGCAGAGTGGCTCCCTGGCTAAC 480
 DB 556 GGAATGTCTCAACCTTATATCTTCTGCTCCAGCAGAGTGGCTCCCTGGCTAAC 615
 QY 481 GGGGCTCTCCGCTCAATACAGCTCTGCTGCTGGCTTGGCATCTGAGCAGCATGGCCA 540
 DB 616 GGGGCTCTCCGCTCAATACAGCTCTGCTGCTGGCTTGGCATCTGAGCAGCATTTGCCA 675
 QY 541 AAGAGTCTTCTTCAAGCATCTGCTCCAGGCTCAAACTTAAACATTAAGTACAGAAG 600
 DB 676 AAGAGTCTTCTTCAAGCATCTGCTCCAGGCTCAAACTTAAACATTAAGTACAGAAG 735
 QY 601 GCACATCATCTGATGCTCCGAGCGCCCTCCAGCAGCAAGTGGCTGTGCTGAGTCA 660
 DB 736 GCACATCATCTGATGCTCCGAGCGCCCTCCAGCAGCAAGTGGCTGTGCTGAGTCA 795
 QY 661 TCAAGGCTGAATACAGGAGTTTATCAACAGCCAGCAAAAATTAAGTGAAGACTTA 720
 DB 796 TCAAGGCTGAATACAGGAGTTTATCAATGATCAATGATGATGATGATGATGATGAT 855
 QY 721 ACAGGCTCCGAGCAAGAACTATCTCATGCAATCAAGTTTACCCAGGCTGAGTGGCT 780
 DB 856 ACAGGCTCCGAGCAAGAACTATCTCATGCAATCAAGTTTACCCAGGCTGAGTGGCT 915
 QY 781 TCAATATGGAATCTTCTGATCAATGATGATGATGATGATGATGATGATGATGAT 840
 DB 916 TCAATATGGAATCTTCTGATCAATGATGATGATGATGATGATGATGATGATGAT 975
 QY 841 CTAGCTATGCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 976 CTAGCTATGCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1035
 QY 901 CCAGAAATACATCCCT 960
 DB 1036 CCAGAAATACATCCCT 1095
 QY 961 AGCTCTTCT 1020
 DB 1096 AGCTCTTCT 1155
 QY 1021 C--CAGAGAAAGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1077
 DB 1156 CAATTAGAAAGAAATTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1215
 QY 1078 GGCAGTGTGAGCTGAGAGAGCGCGCCAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1137
 DB 1216 GGCAGTGTGAGCTGAGAGAGCGCGCCAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1275
 QY 1138 GAGCGGTTGGCTAGCTGAGAGCGCGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
 DB 1276 GAGCGGTTGGCTAGCTGAGAGCGCGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1335

QY 1198 CAGAGGCGCAAGCGGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 DB 1336 CAAGGCGCAAG 1395
 QY 1258 CGGAGCGAG 1317
 DB 1396 CGGAGCGAG 1455
 QY 1318 CTGAAAGCGAG 1377
 DB 1456 CTGAAAGCGAG 1515
 QY 1378 AGGAAAG 1437
 DB 1516 AGGAAAG 1575
 QY 1438 GAGTTAG 1497
 DB 1576 GAGTTAG 1635
 QY 1498 TGTGAG 1557
 DB 1636 TGTGAG 1695
 QY 1558 ATTGCTGAATCAACCACTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1617
 DB 1696 ATTGCTGAATCAACCACTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1755
 QY 1618 CTTATTCAG 1677
 DB 1756 CTTATTCAG 1815
 QY 1678 CATGAG 1737
 DB 1816 CATGAG 1875
 QY 1738 CAGCTCCGAG 1797
 DB 1876 CAGCTCCGAG 1935
 QY 1798 GATGTTTCAACCAAGCTGAG 1857
 DB 1936 GATGTTTCAACCAAGCTGAG 1995
 QY 1858 AAGCAG 1917
 DB 1996 AAGCAG 2055
 QY 1918 GAGTTAG 1940
 DB 2056 GAGTTAG 2078

RESULT 12
 AAK94139
 ID AAK94139 standard; cDNA; 2131 BP.
 XX
 AC AAK94139;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 2646.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000BP-0114089.
 XX

DB 1825 CTTATTCCGAGAAAAACAGTACTCAATGACCAATTAAAAAAAGTTCCAGCAGAACGTTTG 1884
QY 1678 CATAGAGACTCGCTTCTTACCCCTCAAAAGAGCCTTGGAAGCAAGAGCTGGCCCGCAG 1737
DB 1885 CACAGAGATTCACTTGTACCTTAACAAAGCCTTAGAAGCAAAAGAACTACTGCGCAG 1944
QY 1738 CAGCTCCGGAGAGAGCTGACGACAGGTGAGAGAGAGACCAAGTCAAGTGCAGAGATT 1797
DB 1945 CACCTACGAGACCACTGATGATGAGAGAAAGAACTGATCAAACTACAGAGATT 2004
QY 1798 GATGTTTCAACACAGCTGAAGAGATGATCATGACAAACAGACCTCCAG 1857
DB 2005 GATATTTTCAATATATACCTGAGAGAACTTAGAGAAATACAAATBAGCAACACTCCAG 2064
QY 1858 AAGCAGAGTCCCTGAGAGCAGCGCACTGAAGCAAAAGAGAGAGAGAGCCTG 1917
DB 2065 AAGCAAAAGTCCATGAGAGCTGAGAGCTGAAACAGAAAGAAAGCAAGAAATCAT 2124
QY 1918 GAGTTAG 1924
DB 2125 GAATTAG 2131

RESULT 13
AAS84762
ID AAS84762 standard; cDNA; 2874 BP.
XX AAS84762;
AC AAS84762;
XX 13-FEB-2002 (first entry)
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #20566.
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX P-PSDB; ABG20575.
XX WPI: 2001-639362/73.
XX DR P-PSDB; ABG20575.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX Claim 1; SEQ ID No 20566; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 other;
Query Match 35.5%; Score 1291.4; DB 23; Length 2874;
Best Local Similarity 76.0%; Pred. No. 3.5e-298;
Matches 1778; Conservative 0; Mismatches 266; Indels 297; Gaps 3;
QY 1303 GCGCCAAAACGGGAACTTGAAGGCAACGCACTTGAATGGGAAGGAAACGGAGACAG 1362
DB 1 GCTCAAAAACGGGAACTTGAAGGCAACGCACTTGAATGGGAAGGAAACGGAGACAG 60
QY 1363 GAATCCTGAATCAGAGAAACAAAGAGCAGAGGAGCACCCTGGTCTGAAGCAGAGAG 1422
DB 61 GAATCCTGAATCAGAGAAACAAAGAGCAGAGGAGCAGTGTGATCTGAGAAAGAGAA 120
QY 1423 AAGACTCTGAGATTGATGATGAAGAGCTCTGAATGACAAAAGCATCCTAGAGAA 1482
DB 121 AAGACTTTGGAATTTGATTAAGAGCTCTGAATGATTAAGAGCATCACTAGAGAGAA 180
QY 1483 CTTGAGATATCAGGTGCTGAGCTGAGCAACCCAGAGCAAAATTTGAGAGCAAGAAC 1542
DB 181 CTTGAGATATCAGATGTGATGATGACCAACCAAGGCAAAATTTGAGAGCAACAA 240
QY 1543 TCTAGAGAGCTAAGATTTGCTGAATCACCACCTTACAGCAGCAGTTGCGAGATCTCAG 1602
DB 241 TCTAGAGAGTTGAGATTTGCTGAATCACCACCTTACAGCAGCAGTTGCGAGATCTCAG 300
QY 1603 CAAATGCTTGAGAGACTTATTTCCAGAGAAACAGATACCTGATCAATTAAGAGCTTGAAGCAAAA 1662
DB 301 CAAATGCTTGAGAGACTTATTTCCAGAGAAACAGATACCTGATCAATTAAGAGCTTGAAGCAAAA 360
QY 1663 CAGAGAAACAGTTTGCATTAAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGCAAAA 1722
DB 361 CAGAGAAACAGTTTGCATTAAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGCAAAA 420
QY 1723 GAGCTGCCCGCGCAGCAGCTCCGCGAGCAGCTGACAGAGTGGAGAGAGACAGAGTCA 1782
DB 421 GAACTAGCTCGCGCAGCAGCTTACAGAGACCACTGATGATGAGAGAGAGAACTAGATCA 480
QY 1783 AAGCTGAGAGATTTGATTTTTCACACACAGCTGAGAGAACTGAGAGATATAGC 1842
DB 481 AACTTACAGAGATTTGATTTTTCACATATACAGCTGAGAGAACTGAGAGATATAGC 540
QY 1843 AAACAGCACTCCAGAAAGCAGAGTCCCTGAGAGCAGCGGACTGAAAGCAGAAAGCAG 1902
DB 541 AAGCAACAACTCCAGAAAGCAGAAAGTCCATGAGAGCTGAAAGCAGAAAGCAGAA 600
QY 1903 GAGAGAGAGAGCTGAGTTAGAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1962
DB 601 GAGAGAGAGAGCTGAGTTAGAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 1963 AAGGACAAAGAGCTGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019
DB 661 AAGGACAAAGAGCTGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 2020 CCCCACGAGAGAGACAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
DB 721 CTCACAG 780
QY 2080 AGAGCAAGCCGAG 2139
DB 781 AAGGCAAAAG 840
QY 2140 GCTAAGCTGGCAACCCAGAGACCTGCTTACCAAGAGAAAGCCGCTTACCATTTCT 2199


```

Db      841  GCTAAGCACAGTCTCCAGGACCTGTCCACGCAAGAAAAAGGTCCACTTACATTTTCT 900
Qy      2200  GCACAGAGAGTGTAAAAGTATATATTAACGAGCGCTGATCCCTTTGATCCAGAAAT 2259
Db      901  GCACAGAGAAATGTAAAAGTGTATTAACGAGCACTGTACCCCTTTGATCCAGAAAC 960
Qy      2260  CACGATGATCAACATCCAGCAGAGATATATGTCATGTGTGATGAAAAGCAGACTGGA 2319
Db      961  CATGATGAATCACTATCCAGCAGAGACATAGTCAATGTGTGATGAAAAGCAGAACTGA 1020
Qy      2320  GAGCCGAGATGTGTGAGAGAGCTGAAAGGAGCGGATGTTCCTCGTGAACCTAT 2379
Db      1021  GAACCCGCTGTGTGAGAGATTAAGAAAGACAGAGGTGTTCCTGTGAACCTAT 1080
Qy      2380  GCAGAAAGATTCAGAAATGAGTTCCCACTCCAGCCAAACAGTGAACCACTGACA 2439
Db      1081  GCAGAGAAATCCAGAAATGAGTTCCCGCTCCAGTGAACCAAGTGAATTCACA 1140
Qy      2440  TCTGCCCCCTCCCAACTGCTCTGCTGAGACCCCTCTCTTTGCTGAGTGAATCT 2499
Db      1141  TCTGCCCCCTCCCAACTGCTCTGCTGAGACCCCTCTCTTTGCTGAGTGAATCT 1200
Qy      2500  TCTGAGCCCTCCCAACCCCAACCACTGAGTCACTTCACTGAGTGGCCAGCAGC 2559
Db      1201  TCGAGCCCTCCCAACCCCAACCACTGAGTCACTTCACTGAGTGGCCAGCAGC 1260
Qy      2560  TCAAAAGAGAGCCAGAAACGAGCACTGAGTGAAGTGGAGTCAAGCTTCTGACC 2619
Db      1261  ACGAATGAGAAACAGAAACGAGTAACTGAGATGAGGAGCCAGCCCTCTCTGACC 1320
Qy      2620  GTACCTGTGTGTGCGCAGTTACGCGCAGAGATCAAGCTTTTCCCGCAGCAGCAGC 2679
Db      1321  GTTCAAGTGTGCGCAGTTAAGGAGAGGTCCGCTTTTCTCAGCCAGCAGCAGCAGC 1380
Qy      2680  TCCCTCCCATCTCCCGTCTGTGCGCAGAGTGAAGAGTGGAGTCAAGGCGCAAGCC 2739
Db      1381  TCCCTCCCGTCTCTGTGCTAAGCAGAGTGAAGAGTGGAGGCGTCAAGGCTCAAGCC 1440
Qy      2740  CTGATCTCCCTGAGAGAGCCAAAAGAACACCACTTAAATTTTAAACAAAGTGCATC 2799
Db      1441  CTAATCTCTTGGAGAGCCAAAAGAACACCACTTAAATTTTAAACAAAGTGCATC 1500
Qy      2800  ACCGTTCTGGAACAGACAGATGTGTGTTGAGAAAGTTCAAGTCAAGAGGTTGG 2859
Db      1501  ACCGTTCTGGAACAGACAGATGTGTGTTGAGAAAGTTCAAGTCAAGAGGTTGG 1543
Qy      2860  TTCCCAAGTCTTACGTGAATCTCAATTTCAAGGCGCGTAAAGAAATCCACAGCATGAT 2919
Db      1544  ----- 1543
Qy      2920  ACTGCCCCCTACTGAAGTCTGTAGTCTAAGAGAGTGGCTTCCCGGCGCGCAAGCCA 2979
Db      1544  ----- 1543
Qy      2980  GCCATTCGCCGAGAGAGTTTATTCATGATGACATACAGAGATTTGAGCAGAGAT 3039
Db      1544  ----- 1543
Qy      3040  TTAACCTTTTCAAGCAAGGAGTGTATGTGTATCAAGAAAGATGTGTGCTGTGACG 3099
Db      1591  TTAACCTTTTCAAGCAAGGAGTGTATGTGTATCAAGAAAGATGTGTGCTGTGACG 1650
Qy      3100  GGAACGGTGGGCGAGCAAGTCCGAGTCTTCCCTTAATGATGAGGCTTAAAGATTTCA 3159
Db      1651  GGAACGGTGGGCGAGCAAGTCCGAGTCTTCCCTTAATGATGAGGCTTAAAGATTTCA 1710
Qy      3160  GAGGCTCTGTGAACCTGTGTGGAAGACAGGATTTAGGAAAAAACTGAATTTGCCAG 3219
Db      1711  GAGGCTCTGTGAACCTGTGTGGAAGACAGGATTTAGGAAAAAACTGAATTTGCCAG 1770
Qy      3220  GTTATTCCTTCTACGCTGTCTGTGTGCGGAAACATCCCTGCGTCTGTGCGAGCTG 3279

```

```

Db      1771  GTTATTCCTTCTACGCTGTCTGTGTGCGGAAACATCCCTGCTGTGAGCTG 1830
Qy      3280  ATTCTGATCCGAAAAAAAGAACCCAGGTGATGTGTGGAAGAGAACTGCAAGCTCGAGGG 3339
Db      1831  ATTCTGATCCGAAAAAAAGAACCC----- 1852
Qy      3340  AAAAAAGCCAGATAGGTGTGTTTCCAGCAAAATTTATGTCAAACTTTAAAGCCCGGACA 3399
Db      1853  ----- 1852
Qy      3400  AGCAAAATCAACCCCACTGAGCTACCCAGACCGCAGTGCAGCAGTGTGCAAGTG 3459
Db      1853  ----- 1866
Qy      3460  ATCGGATGTTCGATTTACACCGCCCAAGATGACGAACTAGCTTTACAGAAAGCTGAG 3519
Db      1867  ATTGGATGTTCGATTTACACCGCCCAAGATGACGAACTAGCTTTACAGAAAGCTGAG 1926
Qy      3520  ATCATCAACGTCCTCAACCAAGAGACCCGAGCTGTGAAAGAGAAAGTCAAGTGGGCAA 3579
Db      1927  ATCATCAACGTCCTCAACCAAGAGACCCGAGCTGTGAAAGAGAAAGTCAAGTGGGCAA 1986
Qy      3580  GTTGGGCTCTTCCCATCCATTTATGTAAGCTGACACAGACATGACCCCGAGCAGCAA 3639
Db      1987  GTGGGCTCTTCCCATCCATTTATGTAAGCTGACACAGACATGACCCCGAGCAGCAA 2046
Qy      3640  T 3640
Db      2047  T 2047

RESULT 14
AAH16578
ID   AAH16578 standard; cDNA; 2131 BP.
XX
XX   AAH16578;
XX
XX   26-JUN-2001 (first entry)
XX
XX   Human cDNA sequence SEQ ID NO:15658.
XX
XX   Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX   Homo sapiens.
XX
XX   EP1074617-A2.
XX
XX   07-FEB-2001.
XX
XX   28-JUL-2000; 2000EP-0116126.
XX
XX   29-JUL-1999; 99JP-0248036.
XX
XX   27-AUG-1999; 99JP-0300253.
XX
XX   11-JAN-2000; 2000JP-0118776.
XX
XX   02-MAY-2000; 2000JP-0183767.
XX
XX   09-JUN-2000; 2000JP-0241899.
XX
XX   (HELI-) HELIX RES INST.
XX
XX   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX   WPI; 2001-318749/34.
XX
XX   Primer sets for synthesizing polynucleotides, particularly the 5602
XX
XX   full-length cDNAs defined in the specification, and for the detection
XX
XX   and/or diagnosis of the abnormality of the proteins encoded by the
XX
XX   full-length cDNAs -
XX
XX   Claim 8; SEQ ID 15658; 2537pp + CD ROM; English.
XX
XX   The present invention describes primer sets for synthesizing 5602
XX
XX   full-length cDNAs defined in the specification. Where a primer set

```

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH13742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

CC Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 other;

Query Match 34.6%; Score 1259.2; DB 22; Length 2131;

Best Local Similarity 81.2%; Pred. No. 1.5e-290;

Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

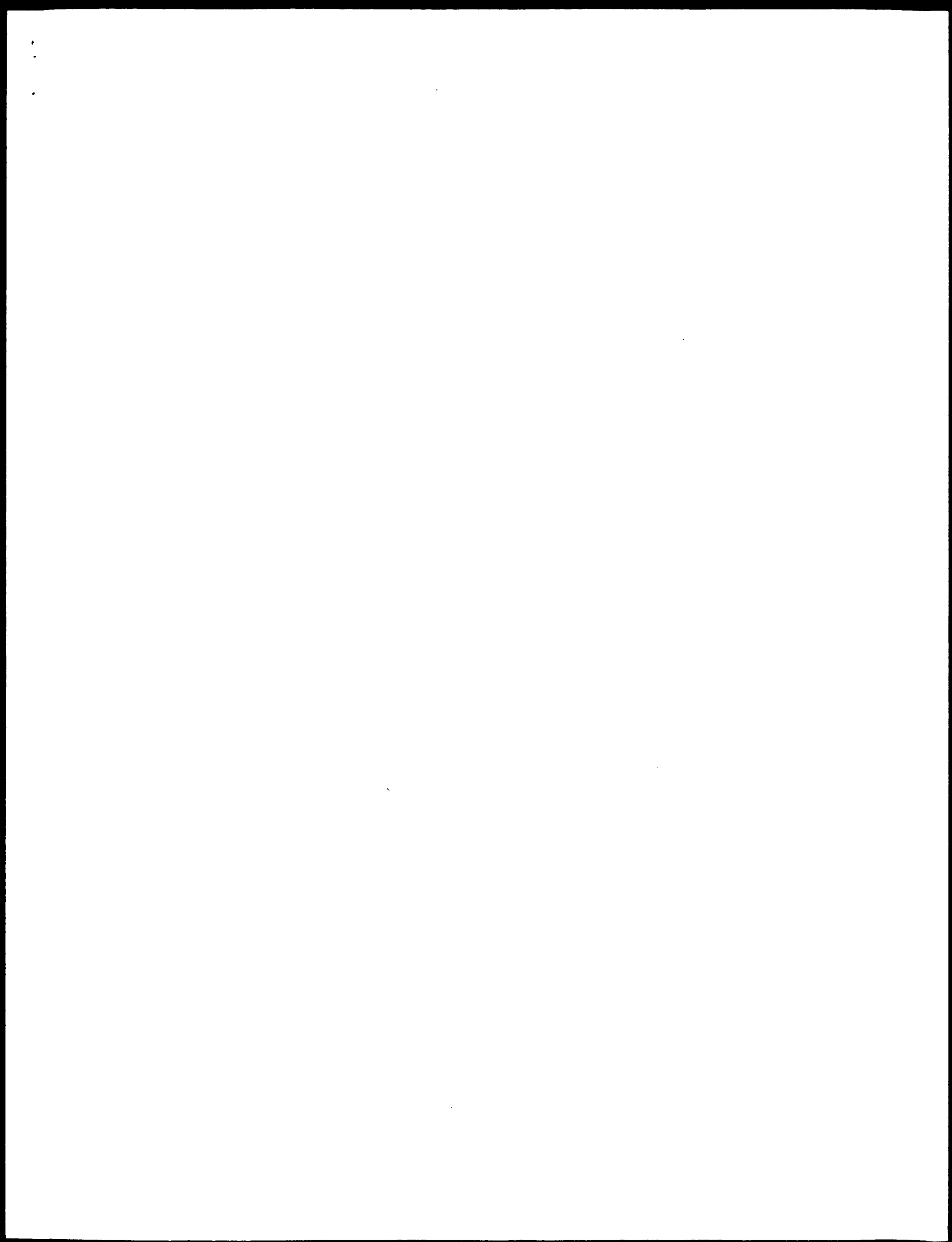
QY 1 AAGGCTCAGTTTCCACACCTTTCGCTGAGCTGATGTCGGGCCCTTAACCTGGAG 60
 DB 347 ATGGCTCAATTTTCAACACCTTTTGGTGGACCTGGATATCTGGGCCATTAACCTGAG 406
 QY 61 GAAAGGCGAAGCATGACGAGCAGTTCCTTAGCTGAAGCCGATAGCGGATTTTACT 120
 DB 407 GAAAGGCGAAGCATGATGACGAGTTCATAGTTTAAAGCCAAATATCTGATTTACT 466
 QY 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTACCTGAGCTTCTTAGACAA 180
 DB 467 GGTGATCAAGCTAGAACTTTTTCATCTGGGTACCTGAGCTTCTTAGACAG 526
 QY 181 ATATGCGCGCTAGCGGACATGATAGATGGAAGGATGATCAAGTGAATTTTCCATA 240
 DB 527 ATATGCGCGCTAGCTGACATGATATGATGGAAGATGATCAAGTGAATTTTCCATA 586
 QY 241 GCCATGAACCTTATCAAACTGAAAGTACAAAGATATCACTCCCTCCACACTTCCCT 300
 DB 587 GCATGAACCTTATCAAACTGAAAGTACAAAGATATCACTCCCTCCACACTTCCCT 646
 QY 301 GTCATGAACAGCAACCACTGCTATTTCCAGTGCACAGCAATTTGGTATAGAGGAT 360
 DB 647 GTCATGAACAGCAACCACTGCTATTTCCAGTGCACAGCAATTTT----- 691
 QY 361 GCTAGATGCAACCACTCAAGCTGTTGCTCCGTGCCAATGGGCTCCATTCCAGTTGT 420
 DB 692 ----- 691
 QY 421 GGAATGTCACACCTTAGATATCTGTCTCCCTCAGACAGAGTGCCTCCCTGGCTAAC 480
 DB 692 -----GCAAGCTGTGCCCCCTGGTAAAC 715
 QY 481 GGGGCTTCCTCCGTCAATACAGCTTGGCTGCTTGGCGATCTGACGCCACATGGCCA 540
 DB 716 GGGGCTTCCTCCGTCAATACAGCTTGGCTGCTTGGCGATCTGACGCCACATGGCCA 775
 QY 541 AAGAGTTCCTCCGTCAAGATGCTGCTCAGGGTCACTTAAACCTAAGTTTACGAAG 600
 DB 776 AAGAGTTCCTCCGTCAAGATGCTGCTCAGGGTCACTTAAACCTAAGTTTACGAAG 835
 QY 601 GCAACAATCATTCGATGCGCAAGCGCCCTCCAGACGAAATGGGCTGTGCTCAGTCA 660
 DB 836 GCAACAATCATTCGATGCGCAAGCGCCCTCCAGACGAAATGGGCTGTGCTCAGTCA 895
 QY 661 TCAAGGCTGAATACAGGCACTTATTAACAGCCACGACAAAATCTATGAGTGACACTTA 720
 DB -----

DB 896 TCAAGCTGAATATACAGGCAATTAATTCATATGATGACAAAATCTATGATGACACTTA 955
 QY 721 ACAGTCCCGACAGCAAACTATTTTCATGCAATCAAGTTTACCCAGGCTCAGTGCT 780
 DB 956 ACAGTCCCGACAGCAAACTATTTTCATGCAATCAAGTTTACCCAGGCTCAGTGCT 1015
 QY 781 TCAATATGAATCTTTTCTGACATTTGATCAAGATGAAAACCTCATGCAAGAAATTTATC 840
 DB 1016 TCAATATGAATCTTTTCTGACATTTGATCAAGATGAAAACCTCATGCAAGAAATTTATC 1075
 QY 841 CTACATATGCAACCTTAATGATGTTGCAATGCTGCTGACGCACTGCCGCCCTGCTGCT 900
 DB 1076 CTGCAATGCAACCTTATGATGTTGCTGCTGACGCACTGCCGCCCTGCTGCT 1135
 QY 901 CCAAAATACATCCCTCTTCTTCTGAGAGAGTTGCTCCGCGAGTGGATGTCCTGAT 960
 DB 1136 CCAAAATACATCCCTCTTCTTCTGAGAGAGTTGCTCCGCGAGTGGATGTCCTGAT 1195
 QY 961 AGCTCTTCTGCTGATCAAGAGCTGCTGAGAGCGCTGCTGAGAGATGAGCAGCAG 1020
 DB 1196 AGCTCAATCTGTGATCAAGAGCTGCTGAGAGCGCTGCTGAGAGATGAGCAGCAG 1255
 QY 1021 C---CAGAGAAAGAACTGCTGATCACTTGAAGATGAAGAGCGGAGAACTTCCAGCGA 1077
 DB 1256 CAATTGAAAGAAATTAATCTGTAACCTTGAAGATGAAGAGCGGAGAACTTGAACGT 1315
 QY 1078 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCCCAAGAGCAG 1137
 DB 1316 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCCCAAGAGCAG 1375
 QY 1138 GAGCGTTGCTGCTGAGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
 DB 1376 GAGCGTTGCTGCTGAGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435
 QY 1198 CAGAGGCGCAAGCGCGAGCTGAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 DB 1436 CAGAGGCGCAAGCGCGAGCTGAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1495
 QY 1258 CCGCAGCGAG 1317
 DB 1496 CAGCAG 1555
 QY 1318 CTGAAAGCGAGCGCAACTTGAATGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
 DB 1556 CTGAAAGCGAGCGCAACTTGAATGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1615
 QY 1378 AGGAACCAAG 1437
 DB 1616 AGGAACCAAG 1675
 QY 1438 GAGTTAGAGCTGATGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
 DB 1676 GAGTTAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1735
 QY 1498 TGTGAGCTGCGCAACCCAGAGGCAAGAAATTTAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 DB 1736 TGTGAGCTGCGCAACCCAGAGGCAAGAAATTTAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
 QY 1558 ATTGCTGAATACACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 DB 1796 ATTGCTGAATACACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
 QY 1618 CTATATTCAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677
 DB 1856 CTATATTCAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1915
 QY 1678 CATAGAGCTGCTTCTTACCTCAAAAGAGCTTGGAGCAAGAGAGAGAGAGAGAGAGAGAG 1737
 DB 1916 CATAGAGCTGCTTCTTACCTCAAAAGAGCTTGGAGCAAGAGAGAGAGAGAGAGAGAGAG 1975
 QY 1738 CAGCTCCGGAG 1797
 DB 1976 CAGCTCCGGAG 2035

QY 2403 GGTTCCCACTCCAGCCAAACCAAGTGCATCTGACATCTGCCCTCGCCCAAACTGGC 2462
 Db 901 GGTTCCCGCTCCAGTAAACCAAGTGCATCTGACATCTGCCCTCGCCCAAACTGGC 960
 QY 2463 TCGGCGTGAAGACCCCGCTCTTGGCAGTGAAGCTTGTGAGCCCTCCAGAACCCCA 2522
 Db 961 CTTCGCTGAGACCCCGCTCTTGGCAGTGAAGCTTGTGAGCCCTCCAGAACCCCA 1020
 QY 2523 CAACTGGGAGACCTTCAGTTCACGTCGCCCAAGCAGACCTTCAAGAACCCGAAACGGA 2582
 Db 1021 TAACTGGGCGACCTTCAGTTCACGTCGCCCAAGCAGACCTTCAAGAACCCGAAACGGA 1080
 QY 2583 CAACTGGGATAGTCGGGCGCTCAGCTTCTCTGACCCGTAAGTGTGCTGGCCAGTTAGC 2642
 Db 1081 TAACTGGGATGATGGGCGACCCGCTCTCTCAGCCGTTCCAGAGCGCGAGTTAG 1140
 QY 2643 GCAGAGATCAGCTTACCCAGCCAGCCAGCCAGCTGCTCCCAATCCCGCTCGGG 2702
 Db 1141 GCAGAGGTCGCTTCTTACCTCCAGCCAGCCAGCCAGCTGCTCCCGCTCTCTGTGCTAGG 1200
 QY 2703 CCAGGCTGAAAAGGTGGAAGGCTACAGCCAGCCAGCTGTATCCCTGGAGAGCCAAAA 2762
 Db 1201 CCAGGCTGAAAAGGTGGAAGGCTACAGCCAGCCAGCTGTATCCCTGGAGAGCCAAAA 1260
 QY 2763 AGACAACTCACTTAATTTTAAACAAAAGTACGTCATCACGCTTCTGGAACAGCAAGACAT 2822
 Db 1261 AGACAACTCACTTAATTTTAAACAAAAGTACGTCATCACGCTTCTGGAACAGCAAGACAT 1320
 QY 2823 GTGCGTGTGGAAGTCAAGGTGAGAAGGTTGTTCCCAAGCTTCTAGTGAAGT 2882
 Db 1321 GTGCGTGTGGAAGTCAAGGTGAGAAGGTTGTTCCCAAGCTTCTAGTGAAGT 1380
 QY 2883 CATTTCAGGCGCCGTAAAGAAATCCACAAGCATCGATCTGGGCCCTAGTGAAGTCTGC 2942
 Db 1381 CATTTCAGGCGCCGTAAAGAAATCCACAAGCATCGATCTGGGCCCTAGTGAAGTCTGC 1440
 QY 2943 TAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGTTCGCGAGAAAGTTAT 3002
 Db 1441 TAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGTTCGCGAGAAAGTTAT 1492
 QY 3003 TGCATGTACACATACGAGAGTTCTGAGCAAGAGATTAACTTTAGCAAGGGATGT 3062
 Db 1493 ----- 1492
 QY 3063 GATTGTGTTACCAAGAAAGATGTGACTGTGACGGGAACGTTGGCGACAAGTCCGG 3122
 Db 1493 ----- 1492
 QY 3123 AGTCTTCCCTTAAGTATGTAGGCTTAAAGATTGAGAGGCTTGAACTGTGGGA 3182
 Db 1493 ----- 1492
 QY 3183 AACAGGAGATTAGAAAAAAACCTGAATTTGCCAGTTATTGCTTCTACGCTGTAC 3242
 Db 1493 -----GAAATTGCCAGTTATTGCTTCTACACACCGCCAC 1527
 QY 3243 TGGTCCCGAACAACCTGAGCTCGTGGCAGCTGATTCGATCCGGAAAAAGAACCC 3302
 Db 1528 CGGCCCCGAGACACTCACTTCGCCCCGTGTCAGCTGATTTTGTATCGAAAAAGAACCC 1587
 QY 3303 AGGTGATGTGGAGAGAACTGCAAGCTCGAGGAAAAAAGCCAGATAGGGTGT 3362
 Db 1588 AGGTGATGTGGAGAGAACTGCAAGCTCGAGGAAAAAAGCCAGATAGGGTGT 1647
 QY 3363 TCCAGCAATTTATGTAATTTTAAAGCCCGGAAACAGCAAAATCACCCCACTGAGCT 3422
 Db 1648 CCCAGCTAATTTATGTAATTTTAAAGCCCGGAAACAGCAAAATCACCCCACTGAGCT 1707
 QY 3423 ACCCAAGACCGCAGTGCAGCAGAGTGCAGAGTGCAGAGTGCAGATTTACACACCGC 3482
 Db 1708 ACCTTAGTCAACAGCATTTAGGCGAGTGTGCCAGGTGATTTGGATGTACACTACCGC 1767
 QY 3483 CCAGAACGATGAGAACTAGCTTTCAGCAAAAGCCAGATCATCACTCTCAACAGGA 3542

Db 1768 GCAGAAATGACATGAGCTGGCTTCAACAAAGGCGAGATCATCAAGTCTCAACAAAGGA 1827
 QY 3543 GACCCGAGCTGTGGAAGAGAGAGTCAAGTGGGCAAGTTGGGCTTTCCATCAATTA 3602
 Db 1828 GACCCCTGACTGTGTGGAAGAGAGTCAATGGAACAAGTGGGCTTTCCATCAATTA 1887
 QY 3603 TGTAAAGCTGACACAGACATGGAACCCAGCCAGCAATGA 3642
 Db 1888 TGTAAAGCTGACACAGACATGGAACCCAGCCAGCAATGA 1927

Search completed: March 14, 2003, 02:10:36
 Job time : 699.685 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:41:40 ; Search time 108.517 Seconds
(without alignments)
10292.544 Million cell updates/sec

Title: US-09-674-237A-2
Perfect score: 3642
Sequence: 1 atggctcagttccacacaccc.....tggaccacccagcagatga 3642
Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	477.4	13.1	2017	4	US-09-404-879A-72
2	332.4	9.1	2873	4	US-08-630-915A-193
3	258	7.1	747	4	US-08-630-915A-193
4	144	4.0	480	4	US-09-404-879A-60
5	144	4.0	531	4	US-09-404-879A-5
6	144	4.0	3489	2	US-08-728-323A-1
7	144	4.0	3489	4	US-09-298-568-1
8	144	4.0	32207	2	US-08-770-379-20
9	144	4.0	32207	4	US-08-757-669A-20
10	144	4.0	32207	4	US-09-230-371A-20
11	121.6	3.3	531	1	US-09-404-879A-4
12	119.6	3.3	9551	2	US-08-056-200-93
13	119.6	3.3	9551	2	US-08-800-644-93
14	119.2	3.3	1926	4	US-09-249-585A-2
15	119.2	3.3	2580	3	US-09-050-863-2
16	119.2	3.3	2580	4	US-09-359-081-2
17	119.2	3.3	5452	2	US-09-130-114-1
18	119.2	3.3	9600	4	US-08-910-647-1
19	119.2	3.3	9600	4	US-09-620-925-1
20	119.2	3.3	10596	1	US-07-884-811-15
21	119.2	3.3	10596	1	US-07-885-971-15
22	119.2	3.3	10596	1	US-08-087-783A-15
23	119.2	3.3	10596	1	US-08-194-088B-15
24	119.2	3.3	10596	2	US-08-194-087-15
25	119.2	3.3	10596	5	PCT-US93-04648-15
26	107	2.9	7218	1	US-08-232-463-14
27	95.8	2.6	16442	3	US-08-781-891-208

28	87.4	2.4	397	3	US-09-253-691-3	Sequence 3, Appl
29	87.4	2.4	51259	3	US-08-781-891-209	Sequence 209, App
30	86.2	2.4	543	6	5273901-6	Patent No. 5273901
31	80.6	2.2	533	6	5482709-5	Patent No. 5482709
32	80	2.2	381	4	US-09-404-879A-61	Sequence 61, Appl
33	80	2.2	2301	1	US-08-306-691B-23	Sequence 23, Appl
34	80	2.2	2301	5	PCT-US93-06251-78	Sequence 78, Appl
35	79.8	2.2	234	1	US-08-469-802B-3	Sequence 3, Appl
36	79.8	2.2	234	4	US-08-267-803B-3	Sequence 3, Appl
37	77.6	2.1	477	4	US-09-135-994-1	Sequence 1, Appl
38	73.6	2.0	2338	1	US-08-425-069-1	Sequence 1, Appl
39	73.6	2.0	2338	2	US-08-317-844B-1	Sequence 1, Appl
40	73.4	2.0	195	1	US-08-469-802B-2	Sequence 2, Appl
41	73.4	2.0	195	2	US-08-267-803B-2	Sequence 2, Appl
42	73.4	2.0	203	4	US-09-043-303-7	Sequence 7, Appl
43	73.4	2.0	12001	1	US-08-458-568A-11	Sequence 11, Appl
44	71.4	2.0	3337	1	US-08-072-610-1	Sequence 1, Appl
45	71.4	2.0	3337	2	US-08-719-822B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-404-879A-72
; Sequence 72, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-404-879A-72

Query Match 13.1%; Score 477.4; DB 4; Length 2017;
Best Local Similarity 55.1%; Pred. No. 8e-110;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

QY	1	ATGGCTCAGTTTCCACACCTTTGGTGTAGCCTGTGATGTTGGCCATTAAGTGAG	60
DB	42	ATGGCTCAGTTTCCACACCTTTGGTGTAGCCTGTGATGTTGGCCATTAAGTGAG	101
QY	61	GAAAGGGCAAGCATGACAGCAGTTCCTTGCTGAAGCCGATAGGGGATTAATACT	120
DB	102	GAACTTACTTAAGCATGATTAACAGTTGATTAACCTCAACCTTGAGAGGTTACTACA	161
QY	121	GATGATCAAGGAGAGAACTTTTTCATCTGCTTACTCAGCTGCTTGTGACAA	180
DB	162	GATGATCAAGGAGAGAACTTTTTCATCTGCTTACTCAGCTGCTTGTGACAA	221
QY	181	ATATGGGGCTAGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT	240
DB	222	ATATGGGGCTAGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT	281
QY	241	GCGATGAAGCTTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	300
DB	282	GCTATGAAGCTTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	341
QY	301	GTCATGAAGCTTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	357
DB	342	ATCATGAAGCTTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	401
QY	358	ATTCATGAAGCTTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	400

Db 402 ATGCCCATCTGTCATTCATACCCATTCGCTCCATTCGACCTATAGCAACACCTTG 461
Qy 401 -----TGGCTCCATTCAGTTGTGGAATGTCTCCACCTTAGTATCT 444
Db 462 TCTTCTGCTACTTCAGGAGACCATATTCCTCCCTAATGATGCTGCTCCCTAGTCCCT 521
Qy 445 TCTGCTCCCTCAGGAGAGTGGCTCCCTGAGTAAAGGAGGCTCCCTCCATACAGCT 504
Db 522 TCTGTAGTACATCTCTATACCAAAATGGAATGCCATCTCATTCAGCTTTATCCATT 581
Qy 505 CTGCTGCTGTTGGCAGTCCTGAGCCATGCGCAAAAGTTCTCTTCAAGATCT 564
Db 582 CCTATCTCTCTCAACATTCCTCATGACATCTTACAGCTGATGAGGAGATTT 641
Qy 565 GGTCCAGGTTCACATTAACATTAAGTTCAGAGGACATTCATTCGATGTCGACG 624
Db 642 GGTGTGCTAGTATCCAGAAAGCCAGTCTGTGATGATTTAGGATCTAGTCAACT 701
Qy 625 GCCCTCCAG-----CAGCAGATGGGCTGTG 651
Db 702 TCCCTCACTGCTTCCCTCAGGAACTCACTAAGACAGGACCTCAGAGTGGGAGTT 761
Qy 652 CCTAGTATCAGGCTGAATACAGGAGTATTCAACGCCAGCAAACTATGACT 711
Db 762 CCTAGCTTCAGATTAAGATTCGCAAAATTTAATAGTCAAGCAAGGATGAGC 821
Qy 712 GACACTTAACAGTCCCGAGCAAGAACTATTCTCATGCAATCAATTTACCCAGGCT 771
Db 822 GGAATCCTCTCAGTTTTCAGAGTCAAGAAATGCCCTTCTCAGTCAATCTCTCAAACT 881
Qy 772 CAGTGGCTTCATATGAAATCTTTCTGACATTTGATCAAGATGAAACTCATCAGAA 831
Db 882 CAGCTACTACTATTGAGCTCTGCTGACATCGATGTAAGGACAGTGAAGAGCTGAA 941
Qy 832 GAATTTATCCAGTATGACACTAATGATGTCATGTCGTCAGCCACTGCCGCC 891
Db 942 GAATTTATCTGCGATGACACTCATGACATGAGCCAAAGCTGACAGCCATACCTG 1001
Qy 892 GTCTGCTCCAGAAATCATCTCTCTCTTCAAGAGTGTGCTCGGAGTGGAGT 951
Db 1002 ACGTTGCTCCGAGCTTGTCTCTCATCTTTCAGAGGAGAAAGCAAG----- 1050
Qy 952 TCCCTCATAGCTCTTCTCTGTGATCAAGGCTGCTGAGGAGCCGTGTCAGAGAT 1011
Db 1051 -----TTGATCTGTTAATGAACTCTGCTTCATATCAAGAAACCAAGAA 1097
Qy 1012 GAGCAGAGCCAGAGAAAGAACTGCTGTGACATTTGAAGTAAAGAGCGGAGAACTTC 1071
Db 1098 GA---AGAGCTCAGAAAGAACTGCAAGTTTATTTGAGGACAAAGCAAGCACTAT 1154
Qy 1072 GAGGAGGCACTGTGAGCTGTGAGAGAGCGCCAGCGCTCTTGGAGCGAGCGGAAA 1131
Db 1155 GAAAGAGAAACATGAGCTGTGAGAGAGAGCCCAAGTGTGATGAGGAGCAGCAGAG 1214
Qy 1132 GAGCAGAGCGGTTGCTGAGCTGAGAGCGCCGAGCAGAGAGAGAGAAAGCGGAGCGC 1191
Db 1215 GAGGCTGAACGCAAGCCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAACTG 1274
Qy 1192 CAGAGAGAGAGCGCAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251
Db 1275 CAGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
Qy 1252 CTGAG 1311
Db 1335 CTGAG 1394
Qy 1312 CGGAGACTGGAAG 1371
Db 1395 CAGAGAGCTTGAAG 1454
Qy 1372 AATCAG 1431

Db 1455 AGTCAG 1514
Qy 1432 GAGTTGAG 1491
Db 1515 CAGCTGAG 1574
Qy 1492 ATCAGAGTGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
Db 1575 GTCCAAATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
Qy 1552 CTAAAGATTTGAG 1611
Db 1635 CTGAGAAATTAAG 1694
Qy 1612 GGAAGACTTAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
Db 1695 ATCTATCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
Qy 1672 AGTTGATAG 1731
Db 1755 AACACAGCTGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
Qy 1732 CGGAG 1791
Db 1815 TGCCAAAG 1874
Qy 1792 GAGATTTGATTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
Db 1875 GAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1934
Qy 1852 CTCAG 1911
Db 1935 TTAGGCTTTGAACATTTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1994
Qy 1912 AGCTGAG 1934
Db 1995 AGATTAG 2017

RESULT 2
US-08-630-915A-193
; Sequence 193, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 193:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2873 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 US-08-630-915A-193

Query Match 9.1%; Score 332.4; DB 4; Length 2873;
 Best Local Similarity 61.0%; Pred. No. 1.9e-73;
 Matches 631; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTCTCTGACCGCTAGCTAGCTGGCCAGTTAGCGAGATAGCTTTACCCCGCA 2668
 DB 541 CTTTTTCAAACTTAACCTTAATAATCATCATGCGAGAAAAATAGCTTCACTCGAACTG 600
 QY 2669 CAGCCACTGGCTCTCCCATCTCCGCTCCTGGCCAGGCTGAAAAAGTGAAGGCTAC 2728
 DB 601 TGTCCCTCG--ATCTGATCACTATTCATGACAGGACAAAGTGTAGAAAACTTAA 657
 QY 2729 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAAAGACCACTTAAATTTTACAAAA 2788
 DB 658 AAGCAGAGGCTTTGTTCTCTGAGCTGCAAGAAAGATTAACCTTGAACTTCTCAAAAC 717
 QY 2789 GTACGCTACACCTTTCTGGAACAGCAAGCATGTGTGTTTGGAGAAAGTTCAAGTTC 2848
 DB 718 ATACATATTAATCTGTCTTGTGACAGCAAGAAAAATGTGGTGGAGGTGATGAG 777
 QY 2849 AGAAGGTTGTTCCCAAGTCTTAAGTCAATTTCAAGGCTGTAAGAAATCA 2908
 DB 778 GAAGAGATGTTTCCCAATCTTAATGTCAAGATCATTTCTGGAGTGAAGTAAACGGG 837
 QY 2909 CAAGCATGATCTGAGCCCTACTGAAAGCTCTGCTTAAGTAAAGAGATGGCTTCCCGG 2968
 DB 838 AA-----GAACCAAGAGCTTTGTATGACAGCTGTAATTAAGAAACCTACCTCGG 885
 QY 2969 CCGCCAAAGCAGCCATTTCCCGGAGAGAGATTATTTGCGATGTCAATACATGAGAGTTCTG 3028
 DB 886 CAGCTTATTCAGTT-----GGAGAGAAATATATGCACTTTTATCCATATTCAGAGTGTG 939
 QY 3029 AGCAGAGATTAACCTTACAGCAAGGGAGTGTGATGTGTTACCAAGAAAGATGTG 3088
 DB 940 AACCTGAGATTTGACTTTCACAGAAAGGTGAAGAAATATGTGTGACCCAGAAAGATGTAG 999
 QY 3089 ACTGTGAGCGGAGACGGTGGGCGACAAAGTCCGAGTCTTCCCTTCACTATGTAGAGC 3148
 DB 1000 AGTGTGAGAGAGAGATGTGAGATGTGAAGTGAATTTTTCATCAAACTATGTCAAAAC 1059
 QY 3149 TTAAGATTCAGAGGCTGTGAACTGTCTGGGAAAAACAGGAGATTAGAAAAAAACCTG 3208
 DB 1060 CAAAGGATCAAGAGAGATTGTGGAGTGTCTAGCAAGTCTGAGCATCAAAATAAAAACCTG 1119
 QY 3209 AAATTTGCCAGTTATGTCTTCTCAAGCTCTACTGTCTCCGAAACAATCACCTGTGCTC 3268
 DB 1120 AGATGTCTAGTAACTTCAAGCATATGTCTTCTGTGTAACAACCTTAGCTTGTGAC 1179
 QY 3269 CTGGGCGAGCTGATTTGATCCGAAAAAAGAACCCAGGTGATGTGGAGAGAGAACTGC 3328
 DB 1180 CAGGACAGTTAATTAATTAATTTAAAGAAAAATACAAAGTGGTGTGCGAAAGAGATGAC 1239
 QY 3329 AAGCTGAGGAGAAAAAGCCGCAAGATAGGTTTCCAGCAAAATATATGCAAACTTTTAA 3388
 DB 1240 AGGCAAGAGAAAAAGAGAGAGAGAGATGTCTTCTGCAATCATATGTTAACTTTTGG 1299
 QY 3389 GCGCCGGAACAGAGAAATACCCCACTAGCTACCCCAAGACCGAGTGCACGCGAG 3448
 DB 1300 GTCCAAAGTATGAAGAGCAACCTGCTTTTATC-----CTG 1338

QY 3449 TGTGCGAGTATGGGATGTAGATTAACACCCCGCAGAGCATGAGACTAGCCTTCA 3508
 DB 1339 TATGTACAGTATGCTATGTATGACTATGACGAATTAATGAGATGACTCAGTTCT 1398
 QY 3509 GCAAAGCCAGATCATCAAGCTCCTCAACAAGAGAGACCCGAGCTGTGGAAAAAGAGAG 3568
 DB 1399 CCAAGGACACTCATTAATTAATTTGAACAAAGATGATCTGATGTGTGCAAGAGAGA 1458
 QY 3569 TCAGTGGGCAAGTTGGGCTCTTCCCATCAATTATGTAAAGCTGACCAAGACATGAGAC 3628
 DB 1459 TCAAGGGGTGACTGCTGCTTCTTCTTCAAACTAGCTTAAGATGACAGACATCAGATC 1518
 QY 3629 CCAGCCAGCATGA 3642
 DB 1519 CAAGTCAACAGTGA 1532

RESULT 3
 US-08-630-915A-39
 Sequence 39, Application US/08630915A

Patent No. 6309820
 GENERAL INFORMATION:
 APPLICANT: SPARKS, Andrew B.
 APPLICANT: HOPMAN, No. 6309820h
 APPLICANT: KAY, Brian K.
 APPLICANT: FOWLES, Dana M.
 APPLICANT: McCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESSES:
 ADDRESS: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/630,915A
 FILING DATE: 03-APR-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 747 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-630-915A-39

Query Match 7.1%; Score 258; DB 4; Length 747;
 Best Local Similarity 63.9%; Pred. No. 4.1e-55;
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

QY 2989 GGAGAGAGTTATGCGCATGACATACAGAGATTCTGAGCAAGAGATTAACTTT 3048
 DB 115 GGAGAGAGATATATTTGCACTTATTCATATTCAGATGTGGAACCTGAGATTACCTTC 174
 QY 3049 CAGCAAGGAGATGTGATTGTGTTACCAAGAAAGATGTGACTGTGTGACGGAGCGGTG 3108

```

Db      175  ACAGAAAGTGAAGAAATATTTGTGACCCAGAAAGATGAGAGTGAGACAGAACTATT 234
QY      3109  GGGACACAGTCCGAGTCTTCCCTTACTAGTGTGGCTTAAAGTTCAAGGGGCTC 3168
Db      235  GGAAGATGAAGTGAATTTTTCATCAAACTATGTCAAACCAAGATCAAGAGATTTT 294
QY      3169  GGAAGTCTGGGAAACAGGAGTTTAGAAAAAACCTGAATTTGCCAGTTATTGCT 3228
Db      295  GGGAGTCTAGCAGTCTGGAGCATCAATTAATAAAACCTGAGATTTGCTCAGTAACTTCA 354
QY      3229  TTCCAGCTGTACTGTCTCCGCAACAACCTGCTGCTCTGGGAGCTGATTTGATC 3288
Db      355  GCATATGTGTCTGTGTCTGAACCACTTAGCTTGCACAGGACGTTAAATTAATT 414
QY      3289  CGGAAAAAGAACCCAGGTGATGTGGAAAGAGAACTGCAAGCTCAGGAAAAAGCCG 3348
Db      415  CTAAGAAAAAATACAAGTGGGTGTGGCAAGAGAGTTACAGGCGAAGAAAAAGCCG 474
QY      3349  CAGATAGGGTGTTCCTCAGCAAAATTATGTCAACTTTCTAAGCCCGGACAGCAAAATC 3408
Db      475  CAGAAAGATGTGTTCTTCCGACGATGTTAACTTTTGAGTCCAGACAGTGAAGAGCC 534
QY      3409  ACCCCACTGAGTACCCAGACCGCAGTGGAGCCAGACGTTGCTCAGTGTGGATG 3468
Db      535  ACACCTGCTTTTCTCTCT-----GTATGTCAAGTGAATTTCTATG 573
QY      3469  TACGATTACACCGCCAGAAAGATGACGAATAGCTTTCAGCAAAAGCCAGATCATCAC 3528
Db      574  TATGATGTAGCAAAATATGAAGATGAGTCAAGTTTCTTCCAAAGGACAATCATTAAT 633
QY      3529  GTCTCAACAGAGAGACCCGAGCTGTGTGAAAGAGAAAGTCAAGTGGGCAAGTTGGCTC 3588
Db      634  GTATGAAACAAAGATGATCTGATTTGTGTGGAGAGAGATCAACGGGGTGAAGTCTC 693
QY      3589  TTCCATTCATTTATGTAAGCTGACACAGACATGACCCAGCCAGCATG 3642
Db      694  TTTCCTCAAACTAGTATGAATGACGACAGCTCAGATCCAGTCAACAGTGA 747

```

```

RESULT 4
US-09-404-879A-60
; Sequence 60, Application US/09404879A
; Patient No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404, 879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-60

```

```

Query Match      4.0%; Score 144; DB 4; Length 480;
Best Local Similarity 62.1%; Pred. No. 1.1e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```

```

QY      1  ATGGCTCAGTTTCCCAACCTTTGCGTGTAGCTGTGAGGCTTGGAG 60
Db      14  ATGGCTCAGTTTCCCAACCTTTGCGTGTAGCTGTGAGGCTTGGAG 73
QY      61  GAAAGGGCAAGCATGACGACGATGCTTCTTGAAGGAGATGAGGAGTTTACT 120
Db      74  GAAAGCTACTAAGCATGATTAACAGTTTGTATTAACCTCAAACTTCAAGAGTTTACT 133

```

```

QY      121  GGTGATCAAGCGAGAACTTTTTTCCATCTGGTTACTAGCTGTCTTACCA 180
Db      124  GGTGATCAAGCCCGTACTTTTTCCTACAGTGTGCGGCGCGGTTTGTAGCGAA 193
QY      181  ATATGGGGCTAGGGAGCATGAATAGATGGAAGATGATCAAGTGAATTTCCATA 240
Db      194  ATATGGGGCTTATCAGATCTGAACAGAGATGGAAGATGGAACGCAAGATTTCTATA 253
QY      241  GGCATGAAGCTTATCAACTGAAGCTACCAAGATATGAGTCCCTCCACACTTCCCT 300
Db      254  GGTATGAAGCTCATCAAGTTAAAGTTGAGGCGCAACAGCTGCTGTATGTCCTCCT 313
QY      301  GTATGAAGAACAGCAACGATGCTATTTCCAGTGCACAGATTTGGTATAGAGGAT 360
Db      314  ATCATGAAGAACACCC-----CTATGTTCTTCCACTAATCTGTGCTTTGGGATG 367
QY      361  GCTAGCATGCACACCACTCAGCTGTTGCTCTGCTGCC 398
Db      368  GGAAGCATGCCCATCTGTCTCATTCATCAGCATTTGCC 405

```

```

RESULT 5
US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patient No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404, 879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

```

```

Query Match      4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 1.2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```

```

QY      1  ATGGCTCAGTTTCCCAACCTTTGCGTGTAGCTGTGAGGCTTGGAG 60
Db      42  ATGGCTCAGTTTCCCAACCTTTGCGTGTAGCTGTGAGGCTTGGAG 101
QY      61  GAAAGGGCAAGCATGACGACGATGCTTGTAGCTGAAGCCGATAGCGGATTTACT 120
Db      102  GAACGTACTAAGCATGATTAACAGTTTGATTAACCTTGAAGAGTTTACTATA 161
QY      121  GGTGATCAAGCGAGAACTTTTTTCCATCTGGTTACTAGCTGTCTTACCA 180
Db      162  GGTGATCAAGCCCGTACTTTTTCCTACAGTGTGCGGCGCGGTTTGTAGCGAA 221
QY      181  ATATGGGGCTAGGGAGCATGAATAGATGGAAGATGATCAAGTGAATTTCCATA 240
Db      222  ATATGGGGCTTATCAGATCTGAACAGAGATGGAAGATGGAACGCAAGATTTCTATA 281
QY      241  GGCATGAAGCTTATCAACTGAAGCTACCAAGATATGAGTCCCTCCACACTTCCCT 300
Db      282  GGTATGAAGCTCATCAAGTTAAAGTTGAGGCGCAACAGCTGCTGTATGTCCTCCT 341
QY      301  GTATGAAGAACAGCAACGATGCTATTTCCAGTGCACAGATTTGGTATAGAGGAT 360
Db      342  ATCATGAAGAACACCC-----CTATGTTCTTCCACTAATCTGTGCTTTGGGATG 395
QY      361  GCTAGCATGCACACCACTCAGCTGTTGCTCTGCTGCC 398
Db      368  GGAAGCATGCCCATCTGTCTCATTCATCAGCATTTGCC 433

```


QY 975 GGATCAGAGGCTGCTTGGAGAGCCGCTGCTCAGAGATGAGCAGCAGCCAGAGAAACT 1034
 Db 20419 GCAGAGGCCACACAGAGAGAGCCACAGCAGCAGAGAGCCACACAGAGAGAGCAGCA 20360
 QY 1035 GCTGTGACATTTTGAAGATTAAGAAAGGGGAACTTTCAGCGGAGCAGTGTGAGCTGGA 1094
 Db 20359 GCAGAGGCCACACAGAGAGAGCCACAGCAGCAGAGAGCCACACAGAGAGAGCAGCA 20300
 QY 1095 GAAGCGCGCCGCAAGCGCTTGTGAGCAGCAGCCGCAAGAGCAGAGCGGTGGCTGAGCT 1154
 Db 20299 GCGGAGGCCACACAGAGAGAGCCACAGCAGCAGAGAGCCACACAGAGAGAGCAGCA 20240
 QY 1155 GAGCGCGCCGCAAGCGGAGAGAAAGCGGAGCGCCGCAAGAGCAGAGAGCCGCAAGCGGCA 1214
 Db 20239 GCGGAGGCCACACAGAGAGAGCCACAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 20180
 QY 1215 GCTGAGCTGGAGAGAGAGCTGGAGAAAGCAGCGGAGCTGGAGCGGAGCGAGAGAGGA 1274
 Db 20179 GCAGGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 20120
 QY 1275 GAGAGAGAAAGAGATGAGAGAGCGGAGCGCCGCAAGCGGAACTGGAAAGCGAGGAGA 1334
 Db 20119 GCAGGATGAGCAGCAGAGATGAGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 20060
 QY 1335 ACTTGAATGGAGAACGGAACCGAGAGAGAGAACTCTGAATCAGAGAGAAAGAGAGAGGA 1394
 Db 20059 GCAGGATGAGCAGCAGAGATGAGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 20000
 QY 1395 GGGCACCCTGGCTCTTGAAGGCAAGAGAAAGCTCTGGAGTTTGAAGTGAAGCTTGAA 1454
 Db 19999 GCAGGATGAGCAGCAGAGATGAGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 19940
 QY 1455 TGCAAAAAGCATCAGCTTGAAGAAACTTTCAGGATTCAGGTGTCAGCTGCAACCCA 1514
 Db 19939 GCAGGATGAGCAGCAGAGATGAGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 19880
 QY 1515 GAGCAGAAATTTGAGAGCAGCAAGCTTGAAGAGCTTAAGATTTGTAATTCACCCA 1574
 Db 19879 GCAGCAGCAGAGATGAGCAGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCA 19820
 QY 1575 CTTCACAGCAGCAGTTCAGAGAACTTCTCAGCAAAATGCTTGAAGAATTATTCAGAGAAACA 1634
 Db 19819 GCATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCA 19760
 QY 1635 GATATCTCAGTGAACAGTTAAACAAGTCCAGCAGAAAGTTTGCATAGAGACTCGCTTCT 1694
 Db 19759 GCAGCAGCAGCAGAGATGAAACAGAGCAGCAGAGAGAGCAGCAGAGAGAGCAGCA 19700
 QY 1695 TACCTTCAAAAAGAGCTTGGAAACAAAGAGCTGGCCCGCAGCAGCTCCGGAGCAGCT 1754
 Db 19699 GCAGGAGTTAGAGAGCAGAGCAGAGAGTTAGAGATCAGAGCAGAGAGTTAGAGAGCA 19640
 QY 1755 GAGCAGGTGAGAGAGAGCAGAGTCAAGCTCAGAGAGATTTGTTTCAACAACA 1814
 Db 19639 GCAGCAGAGAGTTAGAGAGCAGAGCAGAGAGTTAGAGAGCAGAGAGAGTTAGAGGA 19580
 QY 1815 GCTGAAGGAATCAGAGAGATTAATAGCAAAACGCAATTCAGAACAGAGAGTCCCTGGA 1874
 Db 19579 GCAGGAGCAGAGATTTAGAGAGCAGAGCAGAGTTAGAGAGCAGAGAGAGAGTTAGA 19520
 QY 1875 GAGAGCGCCGCTGAAGCAGAAAGCAGAGAGAGAGCTGGAATTTAGAGAGCAAA 1934
 Db 19519 GCAGCAGAGCAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGCAGAGAGTTAGA 19460
 QY 1935 GGAAGACGCTCAGAGAGAGATTTAGAGAAAGAGCAAGCAATGGCTTGAGCATGTGACGA 1994
 Db 19459 GCAAGAGCAGAGAGTTAGAGAGCAGAGCAGAGCAAGAGCAAGAAATTAGAGAGGT 19400
 QY 1995 GAGAGCAGAGCAGCGCCCGGAAACCCACAGAGAGAGCAGAGCTGAAGAGAGAAACAG 2054
 Db 19399 GAGAGAGCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTAGAGAGGTGGAAGA 19340

QY 2055 TGTCAAGAGAGAGAGCGGAGAGAGAGCCAGCCGGAATTCAGAACAGAGAGCTG 2114
 Db 19339 GCAGAGAGCAGAGAGTTAGAGAGAGTTGGAAGAGCAGAGAGAGAGAGAGAGT 19280
 QY 2115 GCTTTCCATCCGATCAGAGAGCCAGCTTAAGTGGCCACCAGG 2158
 Db 19279 GGAAGAGCAGAGAGCAGAGAGTTAGAGAGAGTTGGAAGAGAGAGG 19236

RESULT 9

US-08-757-669A-20/C
 ; Sequence 20, Application US/08757669A
 ; Patent No. 6183751

GENERAL INFORMATION:
 APPLICANT: Chang, Yuan
 APPLICANT: Bohenzky, Roy A.
 APPLICANT: Russo, James J.
 APPLICANT: Edelman, Isidore S.
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 NUMBER OF INVENTION: SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/757,669A
 FILING DATE:
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 45185-F
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 32207 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-757-669A-20

Query Match 4.0%; Score 144; DB 4; Length 32207;
 Best Local Similarity 45.1%; Pred. No. 9,3e-26;
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 975 GGATCAGAGGCTGCTTGAAGAGCCGCTGCTCAGAGATGAGCAGCAGCCAGAGAAACT 1034
 Db 20419 GCAGAGGCCACACAGAGAGAGCCACAGCAGCAGAGAGCCACACAGAGAGAGCAGCA 20360
 QY 1035 GCTGTGACATTTTGAAGATTAAGAAAGGGGAACTTTCAGCGGAGCAGTGTGAGCTGGA 1094
 Db 20359 GCAGAGGCCACACAGAGAGAGCCACAGCAGCAGAGAGCCACACAGAGAGAGCAGCA 20300
 QY 1095 GAAGCGCGCCGCAAGCGCTCTTGAAGAGAGAGCCGCAAGAGCAGAGAGCGTTGCTCAGCT 1154
 Db 20299 GCGGAGGCCACACAGAGAGAGCCACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 20240
 QY 1155 GAGCGCGCCGCAAGCAGAGAGAGAAAGCGGAGCGCCGCAAGAGCAGAGAGCGCA 1214
 Db 20239 GCGGAGGCCACACAGAGAGAGCCACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 20180

US-08-056-200-93

Query Match 3.3%; Score 119.6; DB 1; Length 9551;
 Best Local Similarity 44.1%; Pred. No. 6.5e-20;
 Matches 500; Conservative 0; Mismatches 634; Indels 0; Gaps 0;

Qy 975 GGATCAGAGGCTGCTTGAAGAGCCGCTGCTAGAGATGACGACGACCAAGAAACT 1034
 Db 3324 GCAGAGAGAGGCGCGAGAGAGCGGAGACACAGAGAGAGAGCGCGAGCA 3383
 Qy 1035 GCCGTGACATTTGAAGATTAAGAACCGGGAATCTTGAGCGAGCGAGTGTGAGCTGGA 1094
 Db 3384 GCAGCTGAGCGCGAG 3443
 Qy 1095 GAAAGCGCGCGAG 1154
 Db 3444 GAAAGCGCGCGAG 3503
 Qy 1155 GAGAGCGCGCGAG 1214
 Db 3504 GAGAGCGCGAG 3563
 Qy 1215 GCTGAGCTGAG 1274
 Db 3564 GCGGAG 3623
 Qy 1275 GAGAGAGAGAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
 Db 3624 GAG 3683
 Qy 1335 ACTTGAAATGGAAG 1394
 Db 3684 GCTGAG 3743
 Qy 1395 GGGAG 1454
 Db 3744 CGAG 3803
 Qy 1455 TGACAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1514
 Db 3804 GAAAGCGAG 3863
 Qy 1515 GAGGAG 1574
 Db 3864 CGAG 3923
 Qy 1575 CTTCAG 1634
 Db 3924 GCAAG 3983
 Qy 1635 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1694
 Db 3984 GCTGAG 4043
 Qy 1695 TACCTTCAAAAG 1754
 Db 4044 AAG 4103
 Qy 1755 GGAAG 1814
 Db 4104 CGAG 4163
 Qy 1815 GCTGAG 1874
 Db 4164 GCTGAG 4223
 Qy 1875 GCGAG 1934
 Db 4224 CGAG 4283
 Qy 1935 GGAAG 1994
 Db 4284 GCGGCTGAG 4343

Qy 1995 GGAG 2054
 Db 4344 GCAG 4403
 Qy 2055 TGTGAG 2108
 Db 4404 GCTGAG 4457

RESULT 13
 US-08-800-644-93

; Sequence 93 Application US/08800644

; Patent No. 5958752

; GENERAL INFORMATION:

; APPLICANT: Steinert, Peter M.

; APPLICANT: Lee, Seung-Chul

; APPLICANT: Kim, In-Gyu

; APPLICANT: Chung, Soo-Il

; APPLICANT: Park, Sang-Chul

; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and

; NUMBER OF SEQUENCES: 117

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson & Bear

; STREET: 620 Newport Center Drive, Sixteenth Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: U.S.A.

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/800,644

; FILING DATE: 14-FEB-1997

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/056,200

; FILING DATE: 30-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Fedrick, Michael F.

; REGISTRATION NUMBER: 36,799

; REFERENCE/DOCKET NUMBER: NIH054,001A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (714) 760-0404

; TELEFAX: (714) 760-9502

; INFORMATION FOR SEQ ID NO: 93:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9551 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1507..1644

; FEATURE:

; NAME/KEY: Intron

; LOCATION: 1645..2511

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 2512..8070

US-08-800-644-93

Query Match 3.3%; Score 119.6; DB 2; Length 9551;
 Best Local Similarity 44.1%; Pred. No. 6.5e-20;
 Matches 500; Conservative 0; Mismatches 634; Indels 0; Gaps 0;

OY	975	GGATTAGAGGCTGGCTCTGAGAGCGCCGTGCTCAGAGATATGACACACACACAGAAAGAACT	103
Db	3324	GCAGAGAGAGAGCGCCGACGACGACGAGAGAGCGCCGACACACAGAGAGAGAGCGCGAGCA	3383
OY	1035	GCTGTGACATTTGAAGATAAGAACCGGAGAACTTCGACCGAGCAGTGTGGACTGGA	1094
Db	3384	GCAGCTGAGCGCGCAGCAGAGAGAGAGAGCGCCGACGACAGCAAGCTTAGCGCGGACAGAGGA	3443
OY	1095	GAAAGCGCCGCCAACGCTCTTTGGAGCAGCAGCCGAAAGACAGAGCGGTGGCTCAGCT	1154
Db	3444	GGAAGGCGCGCAGCAGCAGCTGAGCGCCGACGACGAGAGAGAGAGCGCCGACACAGCT	3503
OY	1155	GGAAGCGCCCGCAGCAGAGAGAGAAAGACCGGAGCGCCGAGAGCAGAGAGCGCCAACCGGCA	1214
Db	3504	GAGCGCGCAGCAGAGAGAGAGAGAGAGAGCGCCGACGACAGCTTGAAGCGCGACACACACTGAG	3563
OY	1215	GCTGGAAGCTGAGAGAGCAGCTGGAGAAAGCAGCGGAGCTTGAGACCGCAGCAGAGAGAGGA	1274
Db	3584	GCGCGAGCAGCAGCTGAGAGCGCGCAGCAGCTTGAAGCGCGAGCAGCACTTAGCGCGGGA	3623
OY	1275	GAGAGAGAGAGATTCGAGAGAGCGCCGACGCGCCGAAACCGGAACTTGAGAAAGCAGCGCA	1334
Db	3624	GCACACAGCTGAGGCGCAGCAGCAGCAGCTGAGCGCCGACGACAGCTGAGCGCGCAGACACA	3683
OY	1335	ACTTGAATGGGAACGGAACCGGAGACAGCAAACTCTGAAATCGAGGAAACAAGACACAGGA	1394
Db	3684	GCTGAGCGCGCAGCAGCAGCAGCTTGAAGCGCGCAGCAGAGAGAGAGGACGACGACACAACA	3743
OY	1395	GGGCAGCCGTGCTCTTGAAGGCAAGAGAGAGAAAGACTTGAAGTTTGAAGTTGAAGCTCTGAA	1454
Db	3744	CGAGAGGAGAGAGCGCAGCAGCGCGCTGAAAGCGCCGACGACGAGAGAGAGAGCGCGGATTTGGCT	3803
OY	1455	TGACAAAAAGCATAGCTTGAAGAGAAACTTCAGATATCAGGTGTCGACTTGCGAACCCA	1514
Db	3804	GAAAGCGCAGAGAGAGACGAGAGAGGACGACGACGAGAGAGAGCGCCGACAGCACTGAAACG	3863
OY	1515	GAGCGAAGAAATTGAGAGCAGCAAAAGTCTAAGAGCTTAAAGATTGCTGAATATCACCA	1574
Db	3864	GCACCTAGAGAGAGAGAGAGCGCGCAAGCTTGGCTGAAAGCTTCGAGAGAGAGAGAGCGCGGA	3923
OY	1575	CTTACAGCAGCAGTTTGCAGAGATCTCAGCAAAATGCTTTGAAGACTTATTCGAGAGAAACA	1634
Db	3924	GCACAGAGAGAGAGCGCGCAGCAGCACTTAAAGCGGAGCAGAGAGAGAGAGCGCGGAGCAGCG	3983
OY	1635	GATACTCAGTGACCATTTAAACCAAGTCCAGCAGAAACAGTTTGCATAGAGACTCGCTTCT	1694
Db	3984	GCTGAAGCGCCAGAGAGAGAGAGAGGCTCCAGCAGCGGTTGAGAGAGAGCAGCAACAAT	4043
OY	1695	TACCTCAAAAGAGCTTTGGAGCAAAAGAGAGCTGCGCCGACAGCAGCTCCGGAGACACT	1754
Db	4044	AAGACGGAGCAGAGAGAGAGAGGCTCGAGCAGCTGTGAAAGCCGAGAGAGAAACAAGCT	4103
OY	1755	GGAAGAGGTGAGAGAGAGACCAAGTCCAAAGCTGCAGAGAGATTGATGTTTCAACAACA	1814
Db	4104	CGACAGAGAGAGAGAGAGAGCAGCGGCTGAAAGCGCGCAGCAGAGAGAGAGCGCGCATCAGCT	4163
OY	1815	GCTGAAGAACTGAGAGAGATACATAGCAAAACGCAACTCCAGAAACAGAGTCCCTTGA	1874
Db	4164	GCTGAAGCGCAGAGAGAGAGAGCGCCGACGACGCGGTGAAGCGCGAGCAGAGAAAGAGGCT	4223
OY	1875	GGCAGCCCGACTGAAGCAGAAAGACGAGAGAGAGAAAGAGCTTGAGATTAGAGAGCAAAA	1934
Db	4224	CGACACACGACTGAAGCGCAGAGAGAGGTGAGACATCCGAGCAGAGAGAGAGCGCGCAGGA	4283
OY	1935	GGAAGACGCTCAGAGACGAGTTTACGAGAAAGGAGCAAGCAATGGCTGAGACATTTGCAACA	1994
Db	4284	GCGGCTGAAGCGCAGAGAGCGCGGAGAGAGAGCGCCACGAGCTGTGTAAGAGCGAGGA	4343
OY	1995	GGAGAGCAGCCACGCCCCCGGAAACCCACGAGAGAGCAGACTTGAAGAGAGGAGAGCAG	2054
Db	4344	GCAAGAGAGAGAGCGCCACAGCAGCACTGAGAGCGCGAGCAGCAGAGAAAGCGCGGACAGCG	4403
OY	2055	TGTCAGGAAGAGAGCGCGGAAGAGAGAGCCAAAGCCGGAATTCGAAGACAGCA	2108

```

Db      4404   GCTGAAGCGGCAAGGAGGAGGAAAGAGAGGGCTGCAGACGCCGCTTGAAGCCCGAGCA    4457
                                                                    |||
RESULT 14
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/OD905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; NUMBER OF FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match          3.3%; Score 119.2; DB 4; Length 1926;
Best Local Similarity 53.4%; Pred.No. 3.7e-20;
Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY      973   GTGATTCAGAGGCTGCTGTGAGAGAGCCGTCGTCAGAGATTGAGCAGCACCCAGAGAAGAAA    1032
DB       515   GAGGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAG    574
QY      1033   CTGCCTGTGACATTTGAAGATAAAGAGCGGGAGAACTTGACCAGCGACGAGTGTGAGCTG    1092
DB       515   GGGCGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAG    574
QY      1093   GAGAACCGCCCGCAAGCGCTCTTGAGCAGCAGCGCCAAAGAGCAGAGACCGCTTGCTCAG    1152
DB       635   GGGCGAGAGCGAGAGAGGGGCGAGAGAGGAGGAGGAGGCGAGAGGGGCGAGAGGGGCGAGAG    694
QY      1153   CTGGAAGCGGCCGAGCAGAGAGAGAAAGATCGGAGCGCCACAGAGCAGAGGCCAAGCGG    1212
DB       695   CAGGAGGGGCGAGAGCAGAGAGGGGCGAGAGCGAGGAGGCGAGAGCAGAGAGGGGCGAGAG    754
QY      1213   CAGCTGGAAGCTGGAGAACACACTGGAGAGACGCGGAGCTGGAGCGGCGACGAGAGAG    1272
DB       755   GGGCAGAGGCGAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGGGGCGAGAG    814
QY      1273   GAGAGAGAGAAAGAGATGAGAGGCGCGCAGGCCCCGAAAAACGGGAACTGGAAAGCAGCGA    1332
DB       815   GAGGAGGGGCGAGAGAGGGGCGAGAGAGGAGAGGAGGAGGCGAGAGGGGCGAGAGAGGGG    874
QY      1333   CAACTTGAATGGGAACGGAACCGAGAGACAGGAATCTCTGAATCAGAGAACAGAGAGCAG    1392
DB       875   CAGGAGGGGCGAGAGCAGAGAGGGGCGAGAGCGGAGGAGGAGCGAGAGGGGCGAGAGGGGCGAG    934
QY      1393   GAGGCGACCGTGTCCTTGAAGCAAGCAAGAGAAAGACTCTGGAGCTTGGAG    1440
DB       935   GAGCAGAGGAGGAGGCGAGAGCAGAGAGGGGCGAGAGCGAGAGGTTGGAG    982

RESULT 15
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 611411
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; TITLE OF INVENTION: System
```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 3.3%; Score 119.2; DB 3; Length 2580;
Best Local Similarity 53.4%; Pred. No. 4.2e-20;
Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 973 GTGGATCAGAGGCTGCTAGAGAGCCGCTGCTAGAGATGACGACGCCAGAGAAAGAA 1032
DB 898 GAGGAGGGGCGAGGACGAGAGAGGGGCGAGAGGGGCGAGGAGGAGGGGCGAGAG 957
QY 1033 CTGCTGTGACATTGTAAGTAAGAACCGGAGAACTTCGAGCGAGGCACTGTGAGCTG 1092
DB 958 GGGCAGAGAGGGGCGAGAGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1017
QY 1093 GAGAAGCCGCGCAAGCGCTTTGAGCAGCAGCGCAAAAGCAGAGCGGTTGCTCAG 1152
DB 1018 GGGCAGAGAGCGAGAGGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAG 1077
QY 1153 CTGAGCGCGCGGAGCAGAGAGAGAAAGAGCGGCGCCAGAGGAGGAGGAGGAGGAG 1212
DB 1078 CAGAGAGGGGCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137
QY 1213 CAGCTGAGCTGAGAGCAGACTGAGAAAGCAGCGGAGCTGAGCGGAGCGAGAGAGAG 1272
DB 1138 GGGCAGAGAGCGAGAGGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1197
QY 1273 GAGAGAGAGAAAGAGATGAGAGGCGCGAGGCGCAAAACGGGAATCTGAAAAGCGCA 1332
DB 1198 GAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1257
QY 1333 CAATCTGAATGGGAACGGAACCGGAGACAGGAATCTCTGAATCAGAGGAAACAAGAGCAG 1392
DB 1258 CAGAGAGGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1317
QY 1393 GAGGAGCACTGTGCTCTGAAAGCAGAGAGGAAGACTTGAAGTTTGAAG 1440
DB 1318 GAGCAGAGAGAGGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1365

Search completed: March 14, 2003, 11:19:06
Job time : 358.517 secs

APPLICATION NUMBER: US/09/879,957
 FILING DATE: 13-Jun-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/630,915
 FILING DATE: 03-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Mibrook, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 INFORMATION FOR SEQ ID NO: 193:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2873 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 193:
 US-09-879-957-193

Query Match 9.1%; Score 332.4; DB 10; Length 2873;
 Best Local Similarity 61.0%; Pred. No. 4.6e-75;
 Matches 631; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTTCTCTGACCGTACTAGTCTGGCCAGTACGGCAGAGATCAGCTTTACCCCGCA 2668
 DB 541 CTTTTCACAACTTAACTATCAATCAATGAGCAAAATCAGCTTCACTCGCAACG 600
 QY 2669 CAGCCACTGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2728
 DB 601 TGTCCCTCG--ATCTGATACCTATTATGACACGGACAAAGTGATGAAACTTAA 657
 QY 2729 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAGAACCAACCACTTAAATTTTAA 2788
 DB 658 AAGCAGAGGCTTGTCTGAGCTGCAAGAAAGATTAACCTTGAACCTTCAAAAC 717
 QY 2789 GTGACGTCATCAGCTTCTGGAACAGACAGACATGTGTGTTTGAAGAACTTCA 2848
 DB 718 ATGACATTATTAAGTCTTGTGAGCAGCAAGAAATGTGTGAGGTGACATGAG 777
 QY 2849 AGAAGGTTGTTCCCAAGCTTAACTGTAAGTCAATTCAGGAGCCGTAAGAAATCCA 2908
 DB 778 GAAAGAGATGTTCCCAATCTTATGTCAAGATCATTCTTGAGAGTAAATTAACGG 837
 QY 2909 CAAGCATCGATCTGAGCTTAAAGTCTGCTAGTCTTAAAGAGAGTGGCTTCCCG 2968
 DB 838 AA-----GAACAGAAAGCTTGTATGACGTGTAATAAGAAACCTAACCTCG 885
 QY 2969 CCGCAGAGCCATCCCGAGAGAGATTTATGTCATGTACATACATACAGAGATTCT 3028
 DB 886 CAGCTTATTAAGT-----GGAGAAATATATTTGACCTTATCAATTTCAAGTGT 939
 QY 3029 AGAAGAGATTTAACTTCAAGAGAGAGATGATGTTGTTTCAAGAAAGATGTTG 3088
 DB 940 AACCTGAGATTTGACTTCAAGAGAGTGAAGAAATATTTGGTGACCCCAAAAGATG 999
 QY 3089 ACTGTGAGAGGGAAGCGTGGGAGCAAGTCCGAGCTTCCCTTCAATATATGAGGC 3148
 DB 1000 AGTGTGAGAGAGATTTGAGATGAGATGAGATGAGATTTTCCATCAAACTATGTCA 1059
 QY 3149 TTAAGATTCAGAGGCTTGAAGCTGTGGGAAACAGGAGATTTAGAAAAAACTTG 3208
 DB 1060 CAAAGGATTCAGAGAGATTTGGGAGCTGAGCAAGTCTGAGCATCAATTAATAAA 1119
 QY 3209 AAATTTCCGAGTTATGCTTCTTCAAGCTGCTATGCTCCCGAAACATCACTCGGCT 3268
 DB 1120 AGATTGCTCAGTAACTTCAAGCATATGTGCTTCTGTTCTGAAACCACTTAAGCT 1179
 QY 3269 CTGGGAGAGTATTTCTGATCCGAAAGAAACCAAGAGATGATGTGGAGAGAACTGC 3328

DB 1180 CAGACATTTATTTATTTATTTAAAGAAAAATACAAATGGGTGTGCAAGAGATTAC 1239
 QY 3329 AAGCTCAGGAGAAAAAGGCGCAGATAGGTTGTTCCAGCAAAATTAATGTAACCTTAA 1388
 DB 1240 AGGCCAGAGAAAAAGGCGCAGAGATGTTTCTGCGCAGCATGTTAACTTTGG 1299
 QY 3389 GCCCGGAAACAGCAAAATCAACCCCACTAGCTAACCCAGAGCCGAGTCCAGCCAG 3448
 DB 1300 GTCCAAAGTAGAAAGAGCCACACTGCTTTTCATC-----CTG 1338
 QY 3449 TGTCCAGGATATGGGATGTGACATTTACACCGCCAGAAAGATGAGAACTAGCTTCA 3508
 DB 1339 TATGCAAGTATGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
 QY 3509 GCAAGGCGCATATCATCAAGCTCTCAACAGAGAGACCCGAGCTGTGAAAGAGAAAG 3568
 DB 1399 CCAAGGCAACCTATTTATTTATGAAACAAAGTATCTGATTTGTTGGCAGAGAGA 1458
 QY 3569 TCAGTGGCAAGTTGGCTCTTCCCATCCAAATTTATGTAAGCTGACACAGATGACC 3628
 DB 1459 TCAAGGGGTGATGCTGTCTCTTCTTCAAACTAGATGATGATGATGATGATGAT 1518
 QY 3629 CAGCCAGCAATGA 3642
 DB 1519 CAAGTCAACAGTGA 1532

RESULT 3

US-09-764-868-125
 ; Sequence 125, Application US/09764868
 ; Patent No. US20020168711A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OR INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT232
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1510
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 125
 ; LENGTH: 4210
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-868-125

Query Match 9.1%; Score 331.4; DB 9; Length 4210;
 Best Local Similarity 61.0%; Pred. No. 1e-74;
 Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTTCTCTGACCGTACTAGTCTGGCCAGTACGGCAGAGATCAGCTTTACCCCGCA 2668
 DB 918 CTTTTCACAACTTAACTATCAATCAATGAGCAAAATCAGCTTCACTCGCAACG 977
 QY 2669 CAGCCACTGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2728
 DB 978 TGTCCCTCG--ATCTGATACCTATTATGACACGGACAAAGTGATGAAACTTAA 1034
 QY 2729 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAGAACCAACCACTTAAATTTAA 2788
 DB 1035 AAGCAGAGGCTTGTCTGAGACTGCAAGAAAGATTAACCTTGAACCTTCAAAAC 1094
 QY 2789 GTGACGTATCAAGCTTGTGGAACAGACATGTGTGTTTGAAGAACTTAAAGTCT 2848
 DB 1095 ATGACATTATTAAGTCTTGAAGAGCAAGAAATTTGTGTGTTGGGAGGTGATGAG 1154
 QY 2849 AAGAGGTTGTTCCCAAGCTTAACTGTAAGTCAATTTAGGAGCCGCTAAGAAATCCA 2908
 DB 1155 GAAAGAGATGTTTCCCAATCTTATGTCAAGATCATTTCTGCGGAGTGAAGTAAAG 1214
 QY 2909 CAAGCATGATATGAGCCCTTACTGAAAGTCTGCTAAGTAAAGAGATGCTTCCCGG 2968

QY 3589 TTCCCATCCATTATATTAAGCTGACACAGACATGACCCCGCCAGCAATGA 3642
 DB 694 TTCCCTTCAAACTAGCTTAAGATGACGACGACTGAGATCCAAAGTCAACAGTGA 747

RESULT 5
 US-09-864-761-17127
 ; Sequence 17127, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmca-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 17127
 ; LENGTH: 270
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AP000193.1
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7

OTHER INFORMATION: EST HUMAN HIT: AA773263.1, EVALUE 1.00e-112
 ; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
 ; OTHER INFORMATION: NT HIT: g111526214, EVALUE 0.00e+00
 US-09-864-761-17127

Query Match 5.5%; Score 198.8; DB 10; Length 270;
 Best Local Similarity 87.2%; Pred. No. 2.5e-41;
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAGCCAGCTGAGAGCCAGATGCTTGAGAGAGCTGAAGGAAGCAGC 2358
 DB 1 GTGATGAAGCCAGCTGAGAGCCAGATGCTTGAGAGAGCTGAAGGAAGCAGC 60
 QY 2359 GGATGTTCCCTGCAAACTATGACGAAAGATTCCAGAAATGAGTTCCACTCCAGCC 2418
 DB 61 GGGGTGTTCCCTGCAAACTATGACGAAAGATTCCAGAAATGAGTTCCACTCCAGTG 120
 QY 2419 AACCAGTACCATCTGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478
 DB 121 AACCAGTACCATCTGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180
 QY 2479 GCTCCTTGGCAGTACCTCTTGTGAGCCTCCGCAACCCCAACTGCTGCTGAGACCTTC 2538
 DB 181 GCTCCTTGGCAGTACCTCTTGTGAGCCTCCGCAACCCCAACTGCTGCTGAGACCTTC 240
 QY 2539 AGTTCACGT 2548
 DB 241 AGTTCACGT 250

RESULT 6
 US-09-864-761-26948
 ; Sequence 26948, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmca-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO: 26948
 LENGTH: 286
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000311.1
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
 OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.3
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92
 OTHER INFORMATION: EXPRESSED IN HLA, SIGNAL = 0.98
 OTHER INFORMATION: NT HIT: AP114488.1, EVALU0.00e+00
 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALU0.00e+00
 OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALU0.00e+112
 US-09-674-761-26948

Query Match 5.5% Score 198.8; DB 10; Length 286;
 Best Local Similarity 87.2%; Pred. No. 2,6e-41;
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGACG 2358
 DB 1 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGACG 60
 QY 2359 GGAATGCTTCCCTGCAAACTATGACAGAAATGCTTCCCACTCCAGCC 2418
 DB 61 GGGTGTTCCTCGCAAACTATGACAGAAATGCTTCCCACTCCAGTG 120
 QY 2419 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478
 DB 121 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180
 QY 2479 GCTCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCAACAACTGGCGAGACTTC 2538
 DB 181 GCCCCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCAACAACTGGCGAGACTTC 240
 QY 2539 AGTCCACGT 2548
 DB 241 AGTCCACGT 250

RESULT 7
 US-09-674-761-30453
 Sequence 30453, Application US/09864761
 Patent No. US20020048763A1

GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Penn, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aemica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 PRIOR FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO: 30453
 LENGTH: 297
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000117.1
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
 OTHER INFORMATION: EXPRESSED IN HELI, SIGNAL = 2.4
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
 OTHER INFORMATION: NT HIT: AP114487.1, EVALU0.00e+00
 OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALU0.00e+112
 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALU0.00e+45
 US-09-674-761-30453

Query Match 5.5% Score 198.8; DB 10; Length 297;
 Best Local Similarity 87.2%; Pred. No. 2,7e-41;
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGACG 2358
 DB 1 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGACG 60
 QY 2359 GGAATGCTTCCCTGCAAACTATGACAGAAATGCTTCCCACTCCAGCC 2418
 DB 61 GGGTGTTCCTCGCAAACTATGACAGAAATGCTTCCCACTCCAGTG 120
 QY 2419 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478
 DB 121 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180
 QY 2479 GCTCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCAACAACTGGCGAGACTTC 2538
 DB 181 GCCCCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCAACAACTGGCGAGACTTC 240
 QY 2539 AGTCCACGT 2548


```

; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000049.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
US-09-864-761-864

```

```

Query Match      5.0%; Score 180.8; DB 10; Length 487;
Best Local Similarity 84.6%; Pred. No. 1.4e-36;
Matches 203; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```

```

QY 1039 GTGCAATTTGAAGATAAGAGCGGAGAACTTCAGCGAGGAGCTGTGAGCTGAGAAAG 1098
DB 248 GTAAAGCTTTGAAGATAAGAGCGGAGAACTTCAGCTGCAACTGTGAGAAAG 307
QY 1099 CGCCGCAAGGCTCTTGTGAGAGCAGCGCAAGAGAGAGAGCGCTTGTGCTGAGTGGAG 1158
DB 308 CGAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGAGAGAGCGCTTGTGCTGAGTGGAG 367
QY 1159 CGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1218
DB 368 CGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
QY 1219 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1278
DB 428 GAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487

```

```

RESULT 13
US-09-864-761-311
; Sequence 311, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 311
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000193.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-311

```

```

Query Match      4.5%; Score 163; DB 10; Length 475;
Best Local Similarity 80.9%; Pred. No. 5.3e-32;
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

```

QY 2265 TGAGATCACATCCAGCCAGAGAGATATGTCATGTCATGATTAAGCCAGATGAGAGCC 2324
DB 241 TTAGATTAATTAAGCGTGAATCTTTTCCACAGTGATTAAGCCAGATGAGAGCC 300
QY 2325 AGATGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2384
DB 301 CGGCTGCTGTGAGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 2385 AAAGATTCAGAAATGAGGTTCCCACTCCAGCCAAACCAATGACCATGACATCTGC 2444
DB 361 GAAATTCAGAAATGAGGTTCCCGCTCCAGTAAACCACTGACATGATTCACATCTGC 420
QY 2445 CCCTGCCCCCAAACTGCTGCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2499
DB 421 CCCTGCCCCCAAACTGCTGCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475

```

```

RESULT 14
US-09-864-761-13884
; Sequence 13884, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456

```

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 13864
LENGTH: 475
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
US-09-864-761-13884

Query Match 4.5%; Score 163; DB 10; Length 475;
Best Local Similarity 80.9%; Pred. No. 5.3e-32;
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2265 TGAGATCACCATCCAGCCAGAGATATAGTCATGTGATGAAAGCCAGACTGAGAGCC 2324
DB 241 TTAGTTCTACTTAAGCTGTGACTTTTCCACAGGTGATGAAAGCCAAACTGAGAAC 300
QY 2325 AGGATGGCTTGAGGAGAGGTGAAGGAGAGGATGTTCCCTCGCAACTATATGACA 2384
DB 301 CGGCTGGCTTGAGGAGAAATTAAGGAGAGAGAGGTGTTCCCTCGCAACTATATGACA 360
QY 2385 AAGATTCGAGAAATGAGGTTCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2444
DB 361 GAAATTCGAGAAATGAGGTTCCGCTCCAGTGAACAGTACATCTGACATCTGC 420
QY 2445 CCCTGCCCCAAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2499
DB 421 CCCTGCCCCAAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475

RESULT 15

US-09-864-761-17125
Sequence 17125, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aomicca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17125
LENGTH: 180
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000192.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
OTHER INFORMATION: NT HIT: g11526214, EVALUATE 1.00e-84
OTHER INFORMATION: EST HUMAN HIT: BE871545.1, EVALUATE 2.00e-84
OTHER INFORMATION: SWISSPROT HIT: Q50360, EVALUATE 2.20e-02
US-09-864-761-17125

Query Match

4.0%; Score 146.4; DB 10; Length 180;

Best Local Similarity 88.3%; Pred. No. 5.7e-28;
Matches 159; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

QY 347 GTATTAGGAGGATTGCTAGCATGCGACCACTACAGAGTGTTGCTCCTGTCGCAATGGCT 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GATGGAGGATTCGCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTCCCAATGGGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 407 CCATTCCAGTTGTGGAATGTCTCCACCCCTTAGTATCTTCTGTCCCTCCAGCAGCATGTC 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CCATTCCAGTTGTGGAATGTCTCCACCCCTTAGTATCTTCTGTCCCTCCAGCAGCATGTC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 467 CTCCTGCGCTAAGGGGCTCTCCCGTCAATAGACCTCTGCGCTTGGCGCATCTG 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CCCCCCTGGCTAAGGGGCTCTCCCGTCAATAGACCTCTGCGCTTGGCGCATCTG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 14, 2003, 11:37:18
Job time : 695.293 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:40:35 ; Search time 4383.26 Seconds
(without alignments)
13456.658 Million cell updates/sec

Title: US-09-674-237A-2

Perfect score: 3642
Sequence: 1 atggttcagttccacacac.....tggaccacgacgacatga 3642

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estcom: *
17: gb_gse: *
18: em_gse_hum: *
19: em_gse_inv: *
20: em_gse_pin: *
21: em_gse_vrt: *
22: em_gse_fun: *
23: em_gse_mam: *
24: em_gse_mus: *
25: em_gse_other: *
26: em_gse_pro: *
27: em_gse_rnd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496.6	41.1	2079	11	BC013578 Homo sapi
2	1495	41.0	2126	11	BC020269 Homo sapi
3	795.6	21.8	911	14	BO942708 AGENCOURT
4	722.8	19.8	732	14	BM950476 UI-M-EHOP
5	721.2	19.8	738	14	BM944544 UI-M-EHOP
6	658.2	18.1	979	14	BO719508 AGENCOURT

7	656.6	18.0	690	14	BO179493	BO179493	UI-M-EWO-
8	633.6	17.4	650	10	BB656585	BB656585	BB656585
9	606.8	16.7	610	14	BQ443095	BQ443095	UI-M-EVO-
10	602	16.5	603	9	A1594919	A1594919	UI-M-EVO-
11	593.2	16.3	606	10	AM910367	AM910367	UI-M-EVO-
12	563.4	15.5	773	14	BM681943	BM681943	UI-M-EVO-
13	558.8	15.3	639	14	BQ201258	BQ201258	UI-M-EVO-
14	555	15.2	1062	14	BQ898718	BQ898718	AGENCOURT
15	537.8	14.8	724	10	BE373101	BE373101	AGENCOURT
16	533.8	14.7	537	9	AA063751	AA063751	AGENCOURT
17	530.4	14.6	966	12	BG297330	BG297330	AGENCOURT
18	529	14.5	896	14	BO941336	BO941336	AGENCOURT
19	527	14.5	734	9	A1671143	A1671143	AGENCOURT
20	509	14.0	521	9	A1303871	A1303871	AGENCOURT
21	504.6	13.9	875	12	BF122946	BF122946	AGENCOURT
22	489.6	13.4	866	10	BE587712	BE587712	AGENCOURT
23	484.6	13.3	527	10	BB757493	BB757493	AGENCOURT
24	480.8	13.2	600	9	AL134506	AL134506	AGENCOURT
25	479.8	13.2	483	10	BE200514	BE200514	AGENCOURT
26	473.2	13.0	962	12	BG29478	BG29478	AGENCOURT
27	468.4	12.9	641	14	BM728997	BM728997	AGENCOURT
28	468.2	12.9	641	14	BM728997	BM728997	AGENCOURT
29	463.2	12.7	556	9	A1580905	A1580905	AGENCOURT
30	463	12.7	1009	12	BE871545	BE871545	AGENCOURT
31	461.6	12.7	676	10	BB650196	BB650196	AGENCOURT
32	448.4	12.3	510	12	BE952536	BE952536	AGENCOURT
33	444	12.2	1015	14	BO940346	BO940346	AGENCOURT
34	442.2	12.1	447	10	BE655487	BE655487	AGENCOURT
35	441.4	12.1	574	12	BF16839	BF16839	AGENCOURT
36	441	12.1	501	9	A1852070	A1852070	AGENCOURT
37	433.2	11.9	1767	11	BC020921	BC020921	AGENCOURT
38	432.6	11.9	527	9	AL120063	AL120063	AGENCOURT
39	432.4	11.9	443	9	AA492783	AA492783	AGENCOURT
40	422.8	11.6	546	10	AA668653	AA668653	AGENCOURT
41	420.6	11.5	628	14	BO33010	BO33010	AGENCOURT
42	414	11.4	475	10	BE199657	BE199657	AGENCOURT
43	410	11.3	503	12	BF918745	BF918745	AGENCOURT
44	404	11.1	979	14	BQ941411	BQ941411	AGENCOURT
45	403.2	11.1	568	10	AW594325	AW594325	AGENCOURT

ALIGNMENTS

RESULT 1
BC013578
LOCUS
DEFINITION Homo sapiens, similar to intersectin 1 (SH3 domain protein), clone IMAGE:3878242, mRNA.
ACCESSION BC013578
VERSION BC013578.1 GI:15488896
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2079)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcgbs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (ILNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sngc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://image.llnl.gov>
 Series: IRAX Plate: 14 Row: c Column: 5
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4808822
 This clone has the following problem: incomplete processing.

FEATURES

source

1..2079

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3878242"

/tissue_type="Lung, large cell carcinoma"

/clone_id="NIH MGC_68"

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

BASE COUNT 703 a 451 c 492 g 433 t

ORIGIN

Query Match

41.1%; Score 1496.6; DB 11; Length 2079;

Best Local Similarity 86.0%; Pred. No. 0;

Matches 1671; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

QY 1 ATGCTCAGTTCCCAACCTTTGGTGTAGCTGTGATGCTGGCCATTACTGTGGAG 60
 Db 132 ATGCTCAGTTCCCAACCTTTGGTGTAGCTGTGATGCTGGCCATTACTGTGGAG 191
 QY 61 GAAAGGGCAGCATGACACGAGCTTCTAGCTGAAGCCGATAGCGGATTTACT 120
 Db 192 GAAAGGGCAGCATGACACGAGCTTCTAGCTGAAGCCGATAGCGGATTTACT 251
 QY 121 GGTATCAAGCAGGAACTTTTTCATCTGGTTTACTGAGCTGTCTTACGAA 180
 Db 252 GGTATCAAGCAGGAACTTTTTCATCTGGTTTACTGAGCTGTCTTACGAA 311
 QY 181 ATATGGGCGCTAGGAGCATGATATGATGAGATGATGATCAAGTGAATTTCCATA 240
 Db 312 ATATGGGCGCTAGGAGCATGATATGATGAGATGATGATCAAGTGAATTTCCATA 371
 QY 241 GCCATGAAGCTTATCAACTGAAGCTACAAAGATATCAAGTCCCTCCACACTTCCCT 300
 Db 372 GCTATGAAGCTTATCAACTGAAGCTACAAAGATATCAAGTCCCTCCACACTTCCCT 431
 QY 301 GTCATGAAACGACCAAGTGGCTATTTCCAGTCCACAGATTTGGTATAGAGGAT 360
 Db 432 GTCATGAAACGACCAAGTGGCTATTTCCAGTCCACAGATTTGGTATAGAGGAT 491
 QY 361 GCTAGCATGACCAACCATGAGCTGTGCTCTGTGCAATGGGCTCCATTCAGTTGT 420
 Db 492 GCGAGCATGACCAACCGCTTACAGCTGTGCTCTGTGCAATGGGCTCCATTCAGTTGT 551
 QY 421 GGAATGTCTCAACCTTATATCTTTCTGCTCCACAGAGAGTGGCTCCCTGGCTAAC 480
 Db 552 GGAATGTCTCAACCTTATATCTTTCTGCTCCACAGAGAGTGGCTCCCTGGCTAAC 611
 QY 481 GGGGCTCCTCCGTCATACAGCTCTGTGCTGCTGTTGGAGCTCTTGAGCCACATGGCCA 540
 Db 612 GGGGCTCCTCCGTCATACAGCTCTGTGCTGCTGTTGGAGCTCTTGAGCCACATGGCCA 671
 QY 541 AAGAGTCTTCTTCAGCAGATCTGTGCTCAGAGGTACAAATTAACACTAAGTTACAGAG 600
 Db 672 AAGAGTCTTCTTCAGCAGATCTGTGCTCAGAGGTACAAATTAACACTAAGTTACAGAG 731
 QY 601 GCAAAATCATTCGATGTGCGCAGAGCCCTCCACAGAGAGAGTGGGCTGTGCTCACTCA 660
 Db 732 GCAAAATCATTCGATGTGCGCAGAGCCCTCCACAGAGAGAGTGGGCTGTGCTCACTCA 791
 QY 661 TCAAGGCTAAATACAGGCACTTATTAACAAGCCACGACAAACCTATAGATGACACTTA 720
 Db 792 TCAAGGCTAAATACAGGCACTTATTAACAAGCCACGACAAACCTATAGATGACACTTA 851

QY 721 ACAGGTCCCAAGCAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
 Db 852 ACAGGTCCCAAGCAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 911
 QY 781 TCAATATGAATCTTTGACATGATGATCAAGATGAAACCTCACTGACAGAAATTTATC 840
 Db 912 TCAATATGAATCTTTGACATGATGATCAAGATGAAACCTCACTGACAGAAATTTATC 971
 QY 841 CTAGCTATGACCTAATTTGATGTTGCCATGCTGTGATGAGCCACTGCGCCCTGCTGCT 900
 Db 972 CTGCGAATGACCTCATTTGATGATGATGCTGTGCTGCGCAACCACTGCTGCTGCT 1031
 QY 901 CCAGAAATATATCT 960
 Db 1032 CCAGAAATATATCT 1091
 QY 961 AGCTTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTCAGAGATGACAGCAG 1020
 Db 1092 AGCTTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTCAGAGATGACAGCAG 1151
 QY 1021 C--CAGAGAGAACTGCTCTGTGACATTTGAGATTAAGAAACCGGAGAACTTTCAGCGA 1077
 Db 1152 CATTAGAGAAAGAAATTTCTGTAACTTTGAGATTAAGAAACCGGAGAACTTTCAGCGA 1211
 QY 1078 GGCAGTGTGAGCTGAGAGAGGCGCCAGAGCGCTTGTGAGCAGCAGCAGAGAGAG 1137
 Db 1212 GGCAGTGTGAGCTGAGAGAGGCGCCAGAGCGCTTGTGAGCAGCAGCAGAGAGAG 1271
 QY 1138 GAGCGTTTGTCTCAGCTGAGCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
 Db 1272 GAGCGTTTGTCTCAGCTGAGCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1331
 QY 1198 CAGAGGCGCAACCGGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 Db 1332 CAGAGGCGCAACCGGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1391
 QY 1258 CCGCAGCAG 1317
 Db 1392 CCGCAGCAG 1451
 QY 1318 CTGGAAGGCGAGCGCAACTTGAATGGAGACGAAACCGGAGACAGGAACTCTGAATCAG 1377
 Db 1452 CTGGAAGGCGAGCGCAACTTGAATGGAGACGAAACCGGAGACAGGAACTCTGAATCAG 1511
 QY 1378 AGGAACAG 1437
 Db 1512 AGGAACAG 1571
 QY 1438 GAGTTAGAGCTCTGAATGACAAAAAGCATCAGCTAGAGAGAAAACTTCAGGATATCAG 1497
 Db 1572 GAGTTAGAGCTCTGAATGACAAAAAGCATCAGCTAGAGAGAAAACTTCAGGATATCAG 1631
 QY 1498 TGTGACATGCGCAACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 Db 1632 TGTGACATGCGCAACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691
 QY 1558 ATTGCTGAATATCACCACCTTACAGAGCAGTGTGAGAGATTCAGCAATCTTGGAGA 1617
 Db 1692 ATTGCTGAATATCACCACCTTACAGAGCAGTGTGAGAGATTCAGCAATCTTGGAGA 1751
 QY 1618 CTTATTTCCAGAGAAACAATATCTCAGTACCAAGTTTAAACAAGTCCAGCAACAAGTTTG 1677
 Db 1752 CTTATTTCCAGAGAAACAATATCTCAGTACCAAGTTTAAACAAGTCCAGCAACAAGTTTG 1811
 QY 1678 CATGAGACTGCTCTTACCTCAAAAGAGCTTGGAGCAAAAGAGAGAGAGAGAGAGAG 1737
 Db 1812 CATGAGACTGCTCTTACCTCAAAAGAGCTTGGAGCAAAAGAGAGAGAGAGAGAGAG 1791
 QY 1738 CAGCTCCGGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
 Db 1872 CAGCTCCGGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1931
 QY 1798 GATGTTTTCACAAACACACTGAGAGAACTGAGAGAGATACATACCAACAGCAACTCCAG 1857

Db 1932 GATATTTTCAATATATGAGTGAAGAACTAAGAAATACACATTAAGCAACACTCCAG 1991
 Oy 1858 AAGCAGAGTCCCTGGAGGACGCGACTGAAGCAGAAAGAGAGAGAGAGAGCTCG 1917
 Db 1992 AAGCAAAAGTCCATGAGAGCTGACACACTGAACAGAAAGAGAGAGAGATGATATA 2051
 Oy 1918 GAGTTAGAGAGCAAAAGAGAGA 1940
 Db 2052 GAATAGAAAAACAAAAA 2074

RESULT 2

LOCUS BC020269 2126 bp mRNA linear HTC 19-DEC-2001
 DEFINITION Homo sapiens, clone IMAGE:4899011, mRNA.
 ACCESSION BC020269
 VERSION BC020269.1 GI:17939664
 KEYWORDS HTC.
 SOURCE Homo sapiens.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2126)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letecia Heiao, Martin Krzywinski, Keta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zyderdyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 40 Row: n Column: 1
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4504796
 This clone has the following problem: no cloning site /
 microdeletion.

FEATURES

Location/Qualifiers
 1..2126
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4899011"
 /tissue_type="Pancreas, epithelioid carcinoma"
 /clone_lib="NIH MGC 42"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"

BASE COUNT 712 a 462 c 517 g 435 t

ORIGIN

Query Match 41.0%; Score 1495; DB 11; Length 2126;
 Best Local Similarity 85.9%; Pred. No. 0;
 Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;
 Oy 1 ATGGCTAGTTTCCACACCTTTGGTGTAGCTGATGTCTGGGCAATTAAGTGTGAG 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 182 ATGGCTAGTTTCCACACCTTTTGGTGTAGCTGATGTCTGGGCAATTAAGTGTGAG 241
 Oy 61 GAAAGGGCAAGCATGACACAGCACTTCTTACCTGAGCGGATGAGGGATTTATTAAT 120
 Db 242 GAAAGGGCAAGCATGATGACAGCTTCCATAGTTTAAAGCAATATCTGATTCATTA 301
 Oy 121 GGTATCAAGGAGAGAACTTTTTCATCTGGGTTTACTTACGCTGTCTTAAGCA 180
 Db 302 GGTATCAAGGAGAGAACTTTTTCATCTGGGTTTACTTACGCTGTCTTAAGCA 361
 Oy 181 ATATGGGCTAGGAGGAGCATGATTAAGATGAAGATGATCAAGTGAATTTCCATA 240
 Db 362 ATATGGGCTAGGAGGAGCATGATTAAGATGAAGATGATCAAGTGAATTTCCATA 421
 Oy 241 GCAATGAAGCTTATCAACTGAAGCTACAAAGATATGAGTCTCCCTGACACTTCCCT 300
 Db 422 GCTATGAAGCTTATCAACTGAAGCTACAAAGATATGAGTCACTCTGACCTTCCCT 481
 Oy 301 GTCATGAAGCAAGCAAGTGTATTTCCAGTGCACAGACTTTGGTATTAAGAGGATT 360
 Db 482 GTCATGAAGCAAGCAAGTGTATTTCCAGTGCACAGACTTTGGTATTAAGAGGATT 541
 Oy 361 GCTAGCATGCCACACACTCACAGCTGTGCTGCTGCAATGGGCTCATTCAGTTGTT 420
 Db 542 GCACAGATGCCACCGCTTACAGCTGTGCTGCTGCAATGGGATTCATTCAGTTGTT 601
 Oy 421 GGAATGTCTCCACCTTATGATATCTTCTGCTCCCTCCACAGAGTGTCTCCCTGCTAAC 480
 Db 602 GGAATGTCTCCACCTTATGATATCTTCTGCTCCCTCCACAGAGTGTCTCCCTGCTAAC 661
 Oy 481 GGGGCTCTCCCTGCTATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Db 662 GGGGCTCTCCCTGCTATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721
 Oy 541 AAGAGTTCTTCTTACAGAGATCTGTCCAGGATCACAAATTAACATTAAGTTACAGAG 600
 Db 722 AAGAGTTCTTCTTACAGAGATCTGTCCAGGATCACAAATTAACATTAAGTTACAGAG 781
 Oy 601 GCACATATCTGATGTGTGCGAGCGCCCTCCACAGAGAGTGGGCTGTGCTCACTCA 660
 Db 782 GCACATATCTGATGTGTGCGAGGTGCCAGATGCCAGATGGAGAGTGGGCTGTGCTCA 841
 Oy 661 TCAAGCTGAAATATGAGCACTTATTAACAGCCACGACAAATTAAGTGTGCACTTA 720
 Db 842 TCAAGCTGAAATATGAGCACTTATTAACAGCTATGATGATGATGATGATGATGAT 901
 Oy 721 ACAAGTCCCGAGGAGCAATCTCTCATGCAATCAAGTTTACCCAGGCTCACTGCT 780
 Db 902 ACAAGTCCCGAGGAGCAATCTCTCATGCAATCAAGTTTACCCAGGCTCACTGCT 961
 Oy 781 TCAATATGAATCTTCTGACATTTGATCAAGATGAAGAACTCACTGCAAGAAATTAATC 840
 Db 962 TCAATATGAATCTTCTGACATTTGATCAAGATGAAGAACTCACTGCAAGAAATTAATC 1021
 Oy 841 CTAGCTATGACCTTAATGATGTGTCATGTCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 Db 1022 CTAGCTATGACCTTAATGATGTGTCATGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
 Oy 901 CCAGAAATATCT 960
 Db 1082 CCAGAAATATCT 1141
 Oy 961 AGCTCTTCTCTCTGATGACAGAGCTGCTGAGAGCCGTGTGAGAGATGAGCAGAG 1020
 Db 1142 AGCTCTTCTCTGATGACAGAGCTGCTGAGAGCCGTGTGAGAGATGAGCAGAG 1201
 Oy 1021 C---CAGAGAGAACTGCTGTGACATTTGAAGATTAAGAGCGAGAACTTGAAGCA 1077
 Db 1202 CAATTAGAAAGAAATTAACGTGTAACGTTTGAAGTAAGAGCGAGAACTTGAAGCA 1261
 Oy 1078 GCGAGTGTGAGCTGTGAAGAGCGCGCAACGCTCTTGTGAGCAGAGCGGCAAGAGAG 1137
 Db 1262 GCGAGTGTGAGCTGTGAAGAGCGCGCAACGCTCTTGTGAGCAGAGCGGCAAGAGAG 1321

OY	1138	GAGCGGTTGGCTCAGCTGGAGCGCGCCGACGAGAAAGAGAAAGAGCGGAGCGCCAGAG	1197
Db	1322	GAGCGGCTGGCCAGCTGGAGCGCGCGGACGAGAGAGAAAGAGCGGTGAGCGCCAGAG	1381
OY	1198	CAGAGGCCAAGCGGCAAGCTGGAGAAAGCAAGCTGGAGAAAGCAGCGGAGCTGGAG	1257
Db	1382	CAAGAGCGCAAAAGACAACTGGAACTGGAGAAAGCAACTGAAAGACAGCGGAGCTAGAA	1441
OY	1256	CGGACGCGAGAGAGAGAGAGAGAAAGAGATCTGAGAGGCGGAGCGGCGCAAAACGGGAA	1317
Db	1442	CGGACGAGAGAGAGAGAGAGAGAAAGAAATTTGAGAGGCGAGGCTTGCAAAACGGGAA	1501
OY	1318	CTGGAAGGCGACGACGACAACTTGAAATGGAAACGGAAACCGGAGACAGAACTCCGTAATCAG	1377
Db	1502	CTTGAAAGGCCAAGACAACTTGAGTGGAGACGGAACTCGAAGCGACAACTACTAATATCA	1561
OY	1378	AGGAACAAAGAGAGAGAGAGGACACCGTGGTCTTGAAAGCAAGAGAGAAAGCTCTGAGTTT	1437
Db	1562	AGAAACAAAGAAACAAAGAGGACATAGTTGTACTGAAAGCAAGAAAGAAAGACTTTGGAATTT	1621
OY	1438	GAGTTAGAAAGCTGGAATATGACAAAAAGCATCAGCTGGAAGAGAAACCTCAGGATATACGG	1497
Db	1622	GAAATTAAGAGCTCTAATATATATTAAGATCACTAATGAAAGGAACTTCAAGATATTCAG	1681
OY	1498	TGTGCACTGGCAACCCGAGAGGCAAGAAATTGAGAGCAGCAACAGTCTAGAGGCTAAGA	1557
Db	1682	TGTGCAATGACCACCCAAAGGCAAGAAATTGAGAGCACAACAAATCTAGAGATTGAGA	1741
OY	1558	ATTGCTGAAATACCCACTTACAGCAGCAGGTTGCAGGAATCTCAGCAATGCTTGGAGAA	1617
Db	1742	ATTGCGGAATACCCATCTTAACAGCAAAATTAACAGGAATCTCAGCAATGCTTGGAGAA	1801
OY	1618	CTTATTTCCAGAGAAACAGATACTCAGTGACCAAGTTAAACAAGTCCAGAGAAACAGTTTG	1677
Db	1802	CTTATTTCCAGAGAAACAGATACTCAATGACCAATTAAACAAGTTCAGAGAAACAGTTTG	1861
OY	1678	CATAGAGACTGCTTTCTTAACCTCAAAAGAGCTTGGAAACAAAGAGCTGGCCGGCAG	1737
Db	1862	CACAGAGATTCACTTGTTACACTTAAAGAGCCTTGAAAGCAAAAGAACTAAGCTGGCAG	1921
OY	1738	CAGCTCCGGGACAGCTGGAGCAGAGTGGAGAGAGAGACCAAGTCAAAAGTGGAGAGATTT	1797
Db	1922	CACCTACGAGACCAACTGGATGAAGTGGAGAAAGAACTAGATCAAAACTACAGAGATT	1981
OY	1798	GATGTTTTCAACAACACAGCTGAAGAACTGAGAGAGATCTATGCAAAACAGCAACTCCAG	1857
Db	1982	GATATTTTCAATATATAGCTGGAAGAACTAAGAGAAATACAAATATAGCAAACTCCAG	2041
OY	1858	AAGCAAGGTTCCCTGGAGGACGCGACTGAAGCAGAAAGACAGAGAGAGAGAAAGCCTTG	1917
Db	2042	AAGCAAAAGTCCATGGAGGCTGAACGACTGAACAGAAAGAAACAAGAACGAAAGATCATTA	2101
OY	1918	GAGTTAGAGAACCAAAAGGAGAA	1940
Db	2102	GAAATTAAGAAAAAAGAAAAA	2124

[illegible]

JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
rstraus@ucla.edu

Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: Resgen, Invitrogen Corp.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

```
place: LRAM13743   row: ncolumn: 03
High quality sequence start: 10
High quality sequence stop: 564.
```

FEATURES	Location/Qualifiers
source	1., 911

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6314690"
/clone_1lb="NIH MGC 129"
/lab_host="DH10F (phage-resistant)"
/note="Organ: o1factory epithilium; Vector:
pcmv-sport6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned
unidirectionally. Primer: O1a00 dt. Average insert size
2.2 Kb..constructed by ResGen, Invitrogen Corp. Note: this

```

BASE COUNT	271 a	210 c	280 g	150 t
ORIGIN				

Query Match	21.8%	Score 795.6	DB 14	Length 911
Best Local Similarity	98.0%	Pred. No. 56-184		
Matches 890	Conservative	0	Mismatches 9	Indels 9
				Gaps 8

Qy	678	GCAGTATTCAACAGCCACGCAAAACTATATAGTGGACACTTAAACAGTCCCGAGCAAG	737
Db	1	GCAGTATTCAACAGCCACGCAAAACTATATAGTGGACACTTAAACAGTCCCGAGCAAG	60

Accession	Sequence	Position
QY	AACATATTCATGCATCAAGTTTACCCAGGCTACGCTGGCTCAATATGGAAATCTTC	797
738	AACATATTCATGCATCAAGTTTACCCAGGCTACGCTGGCTCAATATGGAAATCTTC	797
Dh	AACTATTTCTATGCATCAAGTTTACCCAGGCTACGCTGGCTCAATATGGAAATCTTC	120
61	AACTATTTCTATGCATCAAGTTTACCCAGGCTACGCTGGCTCAATATGGAAATCTTC	120

Qy	798	TGACATTGATCAAGATGAAAACTCACTGACAGAAATTTATCTTACGTATGACCTAAT	857
Db	121	TGACATTGATCAAGATGAAAACTCACTGACAGAGAA-TTATCTTACGTATGACCTAAT	179

OY 858 TGATGTCGCATGTCTGGTACGCCACTGCCCGCCGTCCTGCTCAGAAATACATCCCTCC 917
DB 100 TTTATTCTCCCATTTCTCTCTCTCTCCCAATCTGCCCCCCCCTCTGCTCTCCAGATATCATCTCTCC 239

918 TTCCCTCAGAGAGATTCCGTCGCGCAGTGGATGTCCCGTCATAAGCTCTTCTGTGGA 977

978 TCAGAGGCTGCCTGAGAGAGCCCGTCGTCAGAGATGAGCAGCAGCCAGAGAACTGCC 1037

1038 TGTGACATTGAGATAGAGCGGAGAACTTCGACCGAGCAGTGTGGAAGCTGGAGAA 1097

1098	GCGCCGCCAAGCGCTTGTGAGCAGCAGCGCAAGAAGCAGGAGCGGTGGCTCAGCTGGA	1157
------	--	------

DB	420	GGGCGCCCAAGCGCTTGGAGACAGCGCGAAAGACAGAGCGGTTGGCTTCAGCTGGG	479
OY	1158	GGCGCGCGAGAGAGAAAGCGCGGAGCGCCAGAGAGCAGAGAGCCAAAGCGCGAGCT	1217

Dd	480	GC	CGCCGAGCAGAGGAAAAGACGCGGAGCGCCACGAGCAGGAAACGCAAGCGCGCAGCT	539
Qy	1218	GG	AGCTGGAAGACAGCTGGAGAGAGCAGCGGGAGCTGGAGGCGGCAGCGAGAGGAGAG	1277

Dd	540	GGAGCTGGAGAAAGCACTTGGAGAAAGCAGCGGGAGCTTGGAGCGGCACGAGAGAGAGAGAG	599
Dy	1278	GAGGAAGGAGATCGAGAGGCGCGAGGCGCAAAACGGGAATTGAAAGGACGACGACACT	1337

Db	600	GAGGAAGGAGATCGAGAGCCGCGAGGCCGCAAAACGGGAATCTGGAAAGGCGACGACAACT	659
Qy	1338	TGAATGGGAAAC-GGAACCGGAGACAGGAACTCTTAATACAGGAAACAAGACGAGAGG	1396
Db	660	TGAATGGGAACGGGAAACCGGAGACGAGAACTCTTAATACAGGAGAAACAAGGCGAGGAGG	719
Qy	1397	GC-AACGTTGGCTCGAAGGGAAGGAAAGACTGTGAGTTTGAGTTAGAAAGCTCTGAA	1454
Db	720	GCCACCCCTGGTCTCGAAGGCAAGGAGAGACTCTGGAGTTTGAGTTACAAAGCTCTTGA	779
Qy	1455	TGACAAAAAGCATCGTAGAAGGAAAACTTAGATATCAAGT-GTGCATCGCAACCC	1513
Db	780	TGACAAAAAGCATCGTAGAAGGAAAACTTCGGATATCAAGTGGTCCATCGGCAACCC	839
Qy	1514	AGA-GGCAAGAAATTGAGAGCAGC-AACAAGTCTAGA-GAGCTAAGAATTGC-TGAATC	1569
Db	840	AGAGGGCAAGAAATTGAGAGCAGCAAGAAAGCTCTAAGAGCTAAAAAATTGCTTGAATC	899
Qy	1570	ACCGACTT 1577	
Db	900	ACCGACTT 907	

RESULT 4	
BM950476	
LOCUS	
DEFINITION	BM950476 732 bp mRNA linear EST 14-MAR-2002
	UI-M-EHOP-buu-o-05-0-UI.r1 NIH-BMAP_EHOP Mus musculus cDNA clone
	IMAGE:5687260 5', mRNA sequence.

VERSION	BM950476.1	GI:19434066
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE	NIH-MGC http://mgc.nci.nih.gov/ , National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
AUTHORS	Contact: Robert Strausberg Ph.D.
TITLE	
JOURNAL	
COMMENT	

Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MCC Clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5
Location/Qualifiers
1. 732
FEATURES
source

```

source
I. .732
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5687260"
/clone_ljb="NIH MAP EHD0"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/notes="Organ: brain; Vector: pYX-asc; Site:1: Ecor I;
Site_2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double-stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is CAGCCACGAC. This library was created for the

```

BASE COUNT	ORIGIN
208 a	University of Iowa Mouse Brain Molecular Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
172 c	
201 g	
148 t	
3	others

Query Match	19.8%;	Score 722.8;	DB 14;	Length 732;
Best Local Similarity	99.3%;	Pred. No. 3.3e-166;		
Matches 724;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

QY	2817	AGACATGTGTGGTTTGGAGAAGTTCAAGGTACAGAGGGTGTGTTCCCAAGCTTTACGT	2876
Db	4	AATNNGTGGTGTGGAGAAGTTCAAGGTACAGAGGGTGTGTTCCCAAGCTTTACGT	63
QY	2877	GAACCTCATTTTCAGAGGCCCTTAAGGAATTCACAAGCATTCGATCTACTGGCCCTTACTGAAG	2936
Db	64	GAACCTCATTTTCAGAGGCCCTTAAGGAATTCACAAGCATTCGATCTACTGGCCCTTACTGAAG	123
QY	2937	TCCGCTAGTGTCTAAAGAGAGTGGCTTCCCGCGCCGCGCAAGCCAGCCATTCCCGGAGAAGA	2996
Db	124	TCCGCTAGTGTCTAAAGAGAGTGGCTTCCCGCGCCGCGCAAGCCAGCCATTCCCGGAGAAGA	183
QY	2997	GTTTATTTCCATGTACACATACGAGAGTTCTGAGCCAAAGAGATTTTAACTTTCCAGCAAGG	3056
Db	184	GTTTATTTCCATGTACACATACGAGAGTTCTGAGCCAAAGAGATTTTAACTTTCCAGCAAGG	243
QY	3057	GGATGTGATTTGTGGTTCACCAAGAAAGATGTGTGACTGTGTGACCGGGAACGTGGGGCGAATA	3116
Db	244	GGATGTGATTTGTGGTTCACCAAGAAAGATGTGTGACTGTGTGACCGGGAACGTGGGGCGAATA	303
QY	3117	GTCGGAGTCTTCCCTTACTATGTGAGGCTTTAAAGATTACAGAGGCTCTGGAATCTGC	3176
Db	304	GTCGGAGTCTTCCCTTACTATGTGAGGCTTTAAAGATTACAGAGGCTCTGGAATCTGC	363
QY	3177	TGGAAAAACAGGGAGTTTGAAGAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTACGC	3236
Db	364	TGGAAAAACAGGGAGTTTGAAGAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTACGC	423
QY	3237	TGCTACTGTGTTCCCAACAACCTCACCCCTGGCTCTCTGGGACGTGATTTCTGATCCGAAAAA	3296
Db	424	TGCTACTGTGTTCCCAACAACCTCACCCCTGGCTCTCTGGGACGTGATTTCTGATCCGAAAAA	483
QY	3297	GAACCCAGGTGATGTGTGGAGAAGAACTGCAAGCTCGAAGGAAAAAGCCCGAGATAGG	3356
Db	484	GAACCCAGGTGATGTGTGGAGAAGAACTGCAAGCTCGAAGGAAAAAGCCCGAGATAGG	543
QY	3357	GTGGTTTCACAGCAATTTATGTCAAACTTTAAGCCCGCGAAACAACAAATCACCCCAAC	3416
Db	544	GTGGTTTCACAGCAATTTATGTCAAACTTTAAGCCCGCGAAACAACAAATCACCCCAAC	603
QY	3417	TGAGTATACCAAGACCGGAGTGCACGACGAGTGTGCCAGTGTGCGGATGTAGATTA	3476
Db	604	TGAGTATACCAAGACCGGAGTGCACGACGAGTGTGCCAGTGTGCGGATGTAGATTA	663
QY	3477	CACGCCCAAGAACGATGACGAATCTAGCCTTAGCAAAAGCCAGATCATCAACGTCTCAA	3536
Db	664	CACGCCCAAGAACGATGACGAATCTAGCCTTAGCAAAAGCCAGATCATCAACGTCTCAA	723
QY	3537	CAAGGAGGA 3545	
Db	724	CAAGGAGGA 732	
RESULT 5			
LOCUS	BM944544		
DEFINITION	UI-M-EHOP-bvfr-j-08-0-UI.r1 NIH BMAP_EHOP Mus musculus cDNA clone		
ACCESSION	BM944544		
VERSION	BM944544.1		
KEYWORDS	EST.		
SOURCE	house mouse.		

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 738)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
Location/Qualifiers
1. 738
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5695975"
/clone_1lb="NIH_BMAP_EH0P"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pyx-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is CAGCCACGAC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institute of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 198 a 213 c 184 g 143 t
ORIGIN
Query Match 19.8%; Score 721.2; DB 14; Length 738;
Best Local Similarity 99.5%; Pred. No. 8.1e-166;
Matches 734; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 2210 CCAGACTGAGAGGCGAGGATGGTGGAGGAGAGCTGAAAGGGAAGCGGATGTTCCC 2169
1 CCAGACTGAGAGGCGAGGATGGTGGAGGAGAGCTGAAAGGGAAGCGGATGTTCCC 60
Db 1 CCAGACTGAGAGGCGAGGATGGTGGAGGAGAGCTGAAAGGGAAGCGGATGTTCCC 60
QY 2270 TGCAGACTGAGAGGCGAGGATGGTGGAGGAGAGCTGAAAGGGAAGCGGATGTTCCC 2429
61 TGCAGACTGAGAGGCGAGGATGGTGGAGGAGAGCTGAAAGGGAAGCGGATGTTCCC 120
Db 1 TGCAGACTGAGAGGCGAGGATGGTGGAGGAGAGCTGAAAGGGAAGCGGATGTTCCC 120
QY 2430 CGATCTGACATCTGCGCCCTGCGCCCAAGCTGCTCGGAGAGCCCTCTCTTGGCC 2489
121 CGATCTGACATCTGCGCCCTGCGCCCAAGCTGCTCGGAGAGCCCTCTCTTGGCC 180
Db 1 CGATCTGACATCTGCGCCCTGCGCCCAAGCTGCTCGGAGAGCCCTCTCTTGGCC 180
QY 2490 AGTGACCTCTTGGAGCCCTTCCACAAACCCCAACACTGGGAGAGCTTCAAGTTCACAGTG 2549
181 AGTGACCTCTTGGAGCCCTTCCACAAACCCCAACACTGGGAGAGCTTCAAGTTCACAGTG 240
Db 1 AGTGACCTCTTGGAGCCCTTCCACAAACCCCAACACTGGGAGAGCTTCAAGTTCACAGTG 240
QY 2550 GCCAGAGAGCTCAAAAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAG 2609
241 GCCAGAGAGCTCAAAAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAG 300
QY 2610 TTCTCTGACGCTACCTGATGCTGCGAGGATGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2669
1 TTCTCTGACGCTACCTGATGCTGCGAGGATGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2669

Db 301 TTCTCTGACGCTACCTGATGCTGCGAGGATGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
QY 2670 AGCCACTGGCTCTCCCATCTCCGCTCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2729
Db 361 AGCCACTGGCTCTCCCATCTCCGCTCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
QY 2730 AGCCAGAGGCTGATCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2789
Db 421 AGCCAGAGGCTGATCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
QY 2790 TGAGCTATCAGCCGCTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2849
Db 481 TGAGCTATCAGCCGCTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 2850 GAAGGTTGTTCCCAAGTCTTACGTAACTATTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2909
Db 541 GAAGGTTGTTCCCAAGTCTTACGTAACTATTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
QY 2910 AAGCATGATCTGCGCCCTTACTGAAAGTCTGATGCTAAAGAGAGTGGCTTCCCGGC 2969
Db 601 AAGCATGATCTGCGCCCTTACTGAAAGTCTGATGCTAAAGAGAGTGGCTTCCCGGC 660
QY 2970 CGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3029
Db 661 CGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 719
QY 3030 GCAAGGAGATTTAACTTT 3047
Db 720 GCAAGGAGATTTAACTTT 737

RESULT 6
LOCUS B0719508 979 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8219793 lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:618111 5', mRNA sequence.
ACCESSION B0719508
VERSION B0719508.1 GI:21858405
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 979)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
Plate: LLM1583 row: C column: 24
High quality sequence stop: 643.
Location/Qualifiers
1. 979
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:618111"
/clone_1lb="lupski_sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCAAGCGTCCG-3' and
5'-GACTAGTTTAAATCGGAGGCGCCGCTT(15)-3'. Size selected >

1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lipski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

BASE COUNT 294 a 247 c 218 g 217 t 3 others

Query Match 18.1%; Score 658.2; DB 14; Length 979;
Best Local Similarity 84.2%; Pred. No. 2.6e-150;
Matches 776; Conservative 0; Mismatches 141; Indels 5; Gaps 3;

QY 291 ACTTCCCTGTCATGAAACAGACACCAAGTGGCTATTTCAGTCACGACATTTGGTAT 350
DB 1 ACTTCCCTGTCATGAAACAGACACCAAGTGGCTATTTCAGTCACGACATTTGGTAT 60
QY 351 AGAGAGGATTTGTAAGATGCCACCACTCAGCTGTTGCTCTGCGCAATGGGCTCAT 410
DB 61 GGGAGGATTCGACGATCCACCGCTTACAGCTGTTGCTCAGTCGCAATGGGCTCAT 120
QY 411 TCCAGTTGTAAGATGTCACACCTTATATCTTCTGCTCCAGCAGAGTGCCTCC 470
DB 121 TCCAGTTGTAAGATGTCACACCTTATATCTTCTGCTCCAGCAGAGTGCCTCC 180
QY 471 CCTGGCTAAAGGGGCTCTCTCCGTCATACAGCCTGCTGCGGTTTGGCATCTGCAC 530
DB 181 CCGGCTAAAGGGGCTCTCTCCGTCATACAGCCTGCTGCGGTTTGGCATCTGCAC 240
QY 531 CACATGGCCAAAGAGTTTCTTCTTCAGCAGATCTGTCAGGGTTCACATTAACCTTA 590
DB 241 CACATGGCCAAAGAGTTTCTTCTTCAGCAGATCTGTCAGGGTTCACATTAACCTTA 300
QY 591 GTTACGAGAGGCACATCATTCATGTCGACAGCGCCCTCCAGCAGAGATGGGCTGT 650
DB 301 ATTACAAAAGGACATCATTTGATGTGGCAGTGTCCACAGTCGACAGATGGGCTGT 360
QY 651 GCTCAGTCATCAAGGCTGAATATACAGGAGTTTATCAACGACGACAAATATATGAG 710
DB 361 TCTCAGTCATCAAGGCTGAATATACAGGAGTTTATCAACGACGACAAATATATGAG 420
QY 711 TGGACACTTAACAGGTGCCAGGAAACTATTCTCATGCAATCAAGTTTATCCAGGC 770
DB 421 TGGACACTTAACAGGTGCCAGGAAACTATTCTCATGCAATCAAGTTTATCCAGGC 480
QY 771 TCGAGTGGCTCAATATGGAATCTTCTGCAATGATCAAGATGAAAACTCACTGACA 830
DB 481 TCGAGTGGCTCAATATGGAATCTTCTGCAATGATCAAGATGAAAACTCACTGACA 540
QY 831 AGAATTATCTAGTATGACCTAAATGATGTGGCATGTCTGATCAAGCACTGCGGC 890
DB 541 GGAATTATCTAGTATGACCTAAATGATGTGGCATGTCTGATCAAGCACTGCGGC 600
QY 891 GGTCTGCTCCAGAAATACATCCCTCTCTTCAAGAAAGTTGCTCCGCGAGTGGAT 950
DB 601 TGTCTGCTCCAGAAATACATCCCTCTCTTCAAGAAAGTTGCTCCGCGAGTGGAT 660
QY 951 GTGCGTATAGTCTTCTCTGTCATGAGAGTGCCTGAGAGACCGGCTGACAGAGA 1010
DB 661 ATCTGTATAGTCTTCTCTGTCATGAGAGTGCCTGAGAGACCGGCTGACAGAGA 720
QY 1011 TGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAG-ATAAGAGCGGAGA 1066
DB 721 TGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAG-ATAAGAGCGGAGA 780
QY 1067 ACTTTCAGAGGAGGAGTGTGAGAGTGAAGAGCGCGCAAGGCTCTTTCGAGCAGCAG 1126
DB 781 ACTTTCAGAGGAGGAGTGTGAGAGTGAAGAGCGCGCAAGGCTCTTTCGAGCAGCAG 839
QY 1127 GCAAGAGCAGAGCGGTTGGCTCAGCTGAGGCGCGCGGAGAGAGAGAAAGAGCGGG 1186
DB 840 GCAAGAGCAGAGCGGTTGGCTCAGCTGAGGCGCGCGGAGAGAGAGAAAGAGAGCT 899
QY 1187 AGCGCCAGAGCAGAGGCGCAA 1208

Db 900 GACCCCAAGAACAGAGCGCAA 921

RESULT 7

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

|||||
B0179493 690 bp mRNA linear EST 30-APR-2002
UI-M-EW0-bw-k-03-0-UI.r1 NIH_EWAP_EW0 Mus musculus cDNA clone
IMAGE:5704058 5', mRNA sequence.
B0179493
B0179493.1 GI:20354985
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 690)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
DNA Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..690
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5704058"
/clone_1ib="NIH BMAP EW0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTCGCTGGA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). "Gene Discovery in the Developing Mouse Nervous
System", supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 185 a 197 c 176 g 129 t 3 others

Query Match 18.0%; Score 656.6; DB 14; Length 690;
Best Local Similarity 98.7%; Pred. No. 5.6e-150;
Matches 681; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 2293 GTCATGTGATGAAAGCAGACTGAGAGCGCCTTGAGAGAGCTGAAGGG 2252
DB 1 GTCATGTGATGAAAG-CAGACTGAGAGCGCCTTGAGAGAGCTGAAGGG 59
QY 2253 AAGAGGGATGTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAATGAGTTCCACT 2412
DB 60 AAGAGGGATGTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAATGAGTTCCACT 119
QY 2413 CCAGCAAAACGATGACGATCTGACATCTGCCCTGCCCAAACTGCTTGGCTGAG 2472

```

Db      120  CCAAGCCAAACGAGGACGATCTGACATCTGCCCCCCCCCAAGCTGGCTGGCGAG 179
      2473  ACCCTCTCTCTTGGCCAGAGACCTCTTCTGAGCCCTCCCAACCCCAACTGGGCA 2532
      180  ACCCTCTCTCTTGGCCAGAGACCTCTTCTGAGCCCTCCCAACCCCAACTGGGCA 239
      2533  GACTTCAGTTCCAGTGGCCCGGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAG 2592
      240  GACTTCAGTTCCAGTGGCCCGGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAG 299
      2593  ACGTGGGCGGCTGACCTCTCTGACCGTCTAGTCTGGCCAGTTAGCGGAGATCA 2652
      300  ACATGGGCGGCTGACCTCTCTGACCGTCTAGTCTGGCCAGTTAGCGGAGATCA 359
      2653  GCCTTTACCCGAGCCAGCAGCAGCTGCTCTCCCATCTCCGCTCGGCGCAGGTGAA 2712
      360  GCCTTTACCCGAGCCAGCAGCAGCTGCTCTCCCATCTCCGCTCGGCGCAGGTGAA 419
      2713  AAGGTGAAAGGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2771
      420  AAGGTGAAAGGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
      2772  CTTAAATTTTAAACAAAGTACGTCATCAGCTTCTGGAACAGACATGTGTGTT 2831
      480  CTTAAATTTTAAACAAAGTACGTCATCAGCTTCTGGAACAGACATGTGTGTT 539
      2832  TGGAGAAAGTTCAGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2891
      540  TGGAGAAAGTTCAGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
      2892  GCGCGTAGGAAATTCAGAGCATGATCTGCGCCCTACTGAAAGTCTCTGTCTTAA 2951
      600  GCGCGTAGGAAATTCAGAGCATGATCTGCGCCCTACTGAAAGTCTCTGTCTTAA 659
      2952  GAGAGTGGCTTCCCGGCGCCGCAAGCCAGC 2981
      660  GAGAGTGGCTTCCCGGCGCCGCAAGCCAGC 689

RESULT 8
BB656585 650 bp mRNA linear EST 26-OCT-2001
LOCUS BB656585 RIKEN full-length enriched, 12 days embryo spinal ganglion
DEFINITION BB656585 Mus musculus cDNA clone D130043B15 5', mRNA sequence.
ACCESSION BB656585
VERSION BB656585.1 GI:16490413
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 650)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Komno, H., Kouda
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
JOURNAL Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

```

```

Normalisation and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Matsubara, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Komno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, T., Aizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
location/Qualifiers
1..650
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D130043B15"
/clone_1b="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site 1: Salt. Site 2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCGCACTGAGTTTCTTTTCTTTTNN 3']. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTAATTAATTAATTCGCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLX I."
BASE COUNT 178 a 180 c 169 g 122 t
ORIGIN
Query Match 17.4%; Score 633.6; DB 10; Length 650;
Best Local Similarity 99.2%; Pred. No. 2,4e-144;
Matches 636; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2205 GGAGAGGTAAAGTGTATATTAACGAGCGCTTACCCCTTTGAATTCAGACTACGA 2264
      1 GGAGAGGTAAAGTGTATATTAACGAGCGCTTACCCCTTTGAATTCAGACTACGA 60
      2265 TGAGTACACCTCCAGCCAGAGATATGATGATGATGAAGCCAGACTGAGAGCC 2324
      61 TGAGTACACCTCCAGCCAGAGATATGATGATGATGAAGCCAGACTGAGAGCC 120
      2325 AGGATGCTTGGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2384
      121 AGGATGCTTGGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
      2385 AAAGATTCAGAAATGAGGTTCCCACTCCAGCCAAACCACTGATGATGATGCG 2444
      181 AAAGATTCAGAAATGAGGTTCCCACTCCAGCCAAACCACTGATGATGATGCG 240
      2445 CCGTCCCGCCAAAGCTGCTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2504

```

Db	241	CCCTGCCCAAACTGGCTCTGGGTGAGACCCCTGCTCTTTCGCAAGTACCTCTTCTGA	300
Qy	2505	GCCTCCCAAAACCCCAAAACCTGGGAGACTTCAGTTCACGCTGGCCAGACGCTCAAA	2566
Db	301	GCCTCCCAAAACCCCAAAACCTGGGAGACTTCAGTTCACGCTGGCCAGACGCTCAAA	360
Qy	2565	CGAAGACCCAGAAACGGACAACTGGGATAGTGGGCGGCTCAGCTCTTCTGTGACGTTACC	2622
Db	361	CGAAGACCCAGAAACGGACAACTGGGATATATGGGCGGCTCAGCTCTTCTGTGACGTTACC	420
Qy	2625	TAGTGTGGCCAGTTACGGCAGAGATCAGCTTTACCCAGCCACAGCCACGTCCTCTC	2688
Db	421	TAGTGTGGCCAGTTACGGCAGAGATCAGCTTTACCCAGCCACAGCCACGTCCTCTC	480
Qy	2685	CCCATCTTCCGCTCTTGGGCCAGGGGTGAAAAAGGTGGAAGGCTCTCAAGCGCAAGCCTGTGA	2744
Db	481	CCCATCTTCCGCTCTTGGGCCAGGGGTGAAAAAGGTGGAAGGCTCTCAAGCGCAAGCCTGTGA	540
Qy	2745	TCCCTGGAGAGCCAAAAAACAACCACTTAATTTTAAACAAAAGAGATCATCCGCT	2800
Db	541	TCCCTGGAGAGCCAAAAAACAACCACTTAATTTTAAACAAAAGAGATCATCCGCT	600
Qy	2805	TCTTGAACAGCAAGACATGTGTGTGTTTGGAGAAGTTCAAG	2845
Db	601	TCTTGAACAGCAAGACATGTGTGTGTTTGGAGAAGTTCAAG	641

RESULT	9
LOCUS	B0443095
DEFINITION	B0443095 610 bp mRNA linear EST 29-MAY-2007
ACCESSION	U1-M4-EV0-bxf-f-j-12-0-U1.r1 NIH_BMP_EVO Mus musculus cDNA clone
VERSION	B0443095.1 GI:21246207 IMAGE:5707499 5', mRNA sequence.

SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 610)
NIH-MGC <http://mgc.nhl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph. D.

Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/BLND at: <http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES	Location/Qualifiers
source	1. .610

FEATURES	Seq primer: pYX-5.
source	Location/Qualifiers
	1..610
	/organism="Mus musculus"
	/strain="C57BL/6"
	/db_xref="taxon:10090"
	/cds="IMAGE:5707499"
	/clone_1db="NIH BMAP EVO"
	/tissue type="whole brain"
	/dev_stage="embryo 15.5 dpc"
	/lab_host="RDH10B (T1 phage resistant)"
	/note="Organ: brain; Vector: pYX-asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then

BASE COUNT
ORIGIN

175 a 147 c 165 g 123 t

cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GCGCGTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene discovery in the developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), 'Hemin Chin, Ph.D., program coordinator."

Query Match	16.7%	Score 606.8;	DB 14;	Length 610;
Best Local Similarity	99.7%;	Pred. No. 8.7e-138;		
Matches 608; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Qy	289	AGGAAATCCCAAGCATTCGATTCGTGGCCCTACTGAAAGCTCTGATGCTTAAAGAGAGT	295
Db	1	AGGAAATCCCAAGCATTCGATTCGTGGCCCTACTGAAAGCTCTGATGCTTAAAGAGAGT	60
Qy	295	GCTTCCCCCGGCGCGCAAGCAGCATTTCCCGGAGAAAGATTATTGCATGTACACATAC	301
Db	61	GCTTCCCCCGGCGCGCAAGCAGCATTTCGCGGAGAAAGATTATTATGTCATGTACACATAC	120

Dy 3019 GAGGTTCTGAGCAAGAGATTAACTTTCAGCAAAGGGATGTGTAATGTGTTACCAAG 3078
Dd 121 GAGAGTCTGAGCAGAGAGATTAACTTTCAGCAAAGGGATGTGTAATGTGTTACCAAG 180

QY 3079 AAAGATGCTACTGCTGACCGGAAACGGTGGCGACAACTCCGAGCTTTCCTTCTAAC 3138

Db 181 AAAGATGGTGA CTGGTGA CGGGA CGGTGG CGACA GTCCTCC CTTCAAC 240

Qy 3139 TATGTAGGCTTAAGATTCCAGAGGSCCTTGGAACTGCTGGAAAAACAGGAGTTTGGGA 3198

Db 241 TATGTAGGCTTAAGATTCTAGAGGCTCTGGAAGCTCTGGGAAACAGGAGTTTAGGA 300

[illegible]

3259 ACCCGAGGCTCCGAGGCGAAGTGTATTCCTGTATTCGGAATAAAGGACCGAGTGTGATGCTGGGA 3318
DB AAAAAAAAAAGGAAAGGCCAGGTAATGCTTCCACGCGTACGTAAGGATCCCGAATCAATC 350

361 ACCCGGCTCTGGGGCAGCTGATTCGATCCGAAAAAGAACCCAGGTGATGGTGGAA 420

3319 GGAGAACTGCAAGCTCGAGGGAAAAAGCCGACAGATAGGGTGTTCCACGCAATTATGTC 3378

Db 421 GGAGAACTCGAAGCTCGAGGGAAAAAGCCGACAGATAGGCTGGTTTCCAGCAATTATGTC 480

QY 3379 AACCTTCTAAGCCCCGAGACAGCAAAATCACCCCACTGAGCTTACCCAGAAGCCGAGTG 3438

Db 481 AAACCTTAAAGCCCGGAAACAGCAAAATCAACCCCAACTGAGCTAACCCAGAGCCGCAATG 540

3439 CAGCCAGCAGTGTGCCAGGTGATCGGGATTCAGATTACACCGCCAGAACGATGACCAA 3498

Db 541 CAGCAGCAGTGTGCCAGTGTATCGGATTAACGATTACACCCGCCAGACGATGACGAA 600

QY 3499 CTAGCCTCA 3508

| | | | | | | | | |

Db 601 CTAGCCTCA 610

RESULT 10

Accession	Gene	Size (bp)	Structure	EST
AF594919	LOCUS	603 bp	mRNA linear	EST 15-MAR
AF594919	LOCUS	603 bp	mRNA linear	EST 15-MAR

DEFINITION
IMAGE:482370 5' similar to TR:O42287 O42287 INTERSECTIN. ;, mRNA
sequence

ACCESSION	Accession:
VERSION	Version:
AI594919	AI594919
AI594919.1	GI:4603967

KEYWORDS
EST.
SOURCE
house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

REFERENCE
1 (bases 1 to 603)
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; M


```

Db      8  CGCTTCCCCCTGATGAAACAGAACCGGTGCTATTTCATGTGACGACGATTGGTA 67
Qy      350 TAGAGGGAATTGCTAGCATGCCACCACTACAGCTGTGCTGCTGCAATGGGCTCCA 409
Db      68  TAGAGGGAATTGCTAGCATGCCACCACTACAGCTGTGCTGCTGCAATGGGCTCCA 127
Qy      410 TTCAGATTGTTGAATGTCTCCACCCCTTAGTATCTTCTGCTCCAGACAGAGTGCCTC 469
Db      128  TTCAGATTGTTGAATGTCTCCACCCCTTAGTATCTTCTGCTCCAGACAGAGTGCCTC 187
Qy      470 CCTGCTAACGGGGGCTCCCTCCGCTATACAGCTGCTGCTGCTTGGCAATCCTGCGAG 529
Db      188  CCTGCTAACGGGGGCTCCCTCCGCTATACAGCTGCTGCTGCTTGGCAATCCTGCGAG 247
Qy      530 CCACATGCGCAAGAGTCTTCTCTCAGACAGATCTGCTGCGAGGTCACAAATTAAACACTA 589
Db      248  CCACATGCGCAAGAGTCTTCTCTCAGACAGATCTGCTGCGAGGTCACAAATTAAACACTA 307
Qy      590 AGTTACAGAAAGGACATCATTTGATGTGCGCAGCGCCCTCCAGACGAGAAATGGGCTG 649
Db      308  AGTTACAGAAAGGACATCATTTGATGTGCGCAGCGCCCTCCAGACGAGAAATGGGCTG 367
Qy      650 TGCTCAGTATTAAGGCTGAATTAACAGGAGTATTCAACAGCCGCAAAACTATATA 709
Db      368  TGCTCAGTATTAAGGCTGAATTAACAGGAGTATTCAACAGCCGCAAAACTATATA 427
Qy      710 GTGACACATTAACAGGCTGCCCGACGAAACTATTTCTCATGCAATCAAGTTTACCCTGAG 769
Db      428  GTGACACATTAACAGGCTGCCCGACGAAACTATTTCTCATGCAATCAAGTTTACCCTGAG 487
Qy      770 CTGAGCTGGCTTCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATCTCATGCTGAG 829
Db      488  CTGAGCTGGCTTCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATCTCATGCTGAG 547
Qy      830 AAGAAATTTATCTTCAAGTATGACCTTAATTTGATTTGCTGCTGCTGCTGCTGCTGCTG 888
Db      548  AAGAAATTTATCTTCAAGTATGACCTTAATTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTG 606

RESULT 12
BM681943.c 773 bp mRNA linear EST 27-FEB-2002
LOCUS      UI-E-EO1-aiw-e-23-0-UI.s1 UI-E-EO1 Homo sapiens cDNA clone
DEFINITION UI-E-EO1-aiw-e-23-0-UI 3', mRNA sequence.
ACCESSION  BM681943
VERSION     BM681943.1 GI:18991839
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 773)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            97044477
JOURNAL    Contact: Soares, MB
MEDLINE    Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.iowa.edu
COMMENT    Tissue Procurement: Dr. Gregg Hageman
            cDNA library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-53, >Poly_A$imple_repeat (matched complement)

```

```

FEATURES
source
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EO1-aiw-e-23-0-UI"
/clone_1b="UI-E-EO1"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EO1 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGGCTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI).
TAG_LIB=UI-E-EO1
TAG_TISSUE=human fetal eye
TAG_SEQ=CGGCTATACC"

BASE COUNT  118 a 182 c 130 g 343 t

ORIGIN
Query Match      15.5%; Score 563.4; DB 14; Length 773;
Best Local Similarity 83.1%; Pred. No. 4,2e-127;
Matches 642; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy      1168 CAGAGAGGAAAGACCGGAGCGCCAGAGAGGAGCCAAAGCGGAGCTGAGCTGGAG 1227
Db      773  CAGAGAGGAAAGAGCGGTGAGCGCCAGAGAGCAAGGCAAAAGACAACTGGAATCTGGAG 714
Qy      1228 AAGCACTGGAAGAACACCGGGAGCTGAGCGCGACGAGAGGAGGAGAGGAGAG 1287
Db      713  AAGCACTGGAAGAACACCGGGAGCTGGAACGCGAGAGAGAGGATGAGAGAGGAAAGAA 654
Qy      1288 ATCAGAGGCGCGAGCGCGGCAAAACGGGAACCTGGAAGGACGACCACTGATGGGAA 1347
Db      653  ATTGAGAGCGCAGAGGCTGCAAAACGGGAACCTGGAAGGACGACCACTGATGGGAA 594
Qy      1348 CGAAACCGAGACAGGAACCTCCTGAATCAGAGGAACAAGAGACGAGGACCGTGTG 1407
Db      593  CGGAATCGAAGCAAGAACTCTAATCAAGAAACAAGAGACGAGCATAGTTGTA 534
Qy      1408 CTGAAGGCAAGAGAGAACTCTGGAAGTTGATTGAAGCTCTGAATGACAAAAGCAT 1467
Db      533  CTGAAGGCAAGAAAGAAAGACTCTTGAATTGAATTAAGCTCTTAATGATTAAGAAACAT 474
Qy      1468 CAGCTAGAAGGAACCTTCAGATATCAGGTGCTGACCTGCACACCGAGGCAAGAAAT 1527
Db      473  CACTAGAAGGGAACCTTCAGATATCAGGTGCTGACCTGCACACCGAGGCAAGAAAT 414
Qy      1528 GAGAGACGAACAAGTCTAGAGAGCTAAGAAATTGCTGAATTCACCCACTTAAGCAGAC 1587
Db      413  GAGAGACGAACAACAATCTAGAGAGTTGAGAAATTCGCAATTCACCCACTTAAGCAGACAA 354
Qy      1588 TTGACGAATCTTCAGCAAAATCTTGAAGACTTATTCAGAGAAACGATATCTGATGAC 1647
Db      353  TTACAGGAATCTTCAGCAAAATCTTGAAGACTTATTCAGAGAAACGATATCTGATGAC 294
Qy      1648 CAGTTAAACAAAGTCCAGAGAAACAGTTTGATATGAGACTCGCTTACTCCCTAAAGA 1707
Db      293  CAATTAAACAAAGTTCCAGAGAAACAGTTTGATATGAGACTCGCTTACTCCCTAAAGA 234

```

QY 1708 GCCTTGAGCAAGAGAGCTGCGCCGAGACAGCTCCGGAGAGCTGAGCGAGTGAG 1767
 Db 233 GCCTTGAGCAAGAGAGCTGCGCCGAGACAGCTCCGGAGAGCTGAGCGAGTGAG 174
 QY 1768 AGAGAGAGCGAGTCAAGAGCTGAGAGATGATGTTTCAACAACAGCTGAGAGAGCTG 1827
 Db 173 AAGAGAACTGATCTCAACAGAGAGATGATGATTTTCAATATATCAGCTGAGAGAGCTG 114
 QY 1828 AGAGAGATGATGAGCAACAGCACTCCAGAGAGAGCTGCGTGGAGAGCGAGCTG 1887
 Db 113 AGAGAAATGATGATGAGCAACAGCACTCCAGAGAGAGAGCTGAGAGAGCTGAGAGCTG 54
 QY 1888 AAGCAGAAAGAGAGAGAGAGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
 Db 53 AAGCAGAAAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1

RESULT 13
 BQ201258/c 639 bp mRNA linear EST 02-MAY-2002
 LOCUS BQ201258
 DEFINITION UI-R-DQ1-clt-1-21-0-UI.s1 UI-R-DQ1 Rattus norvegicus cDNA clone
 VERSION BQ201258
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 639)
 Bernaldo, M.F., Lemmon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: meoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 normalized rat cell line R3327-5a library cDNA library Preparation:
 M.B. Soares Lab Clone distribution: clones will be available
 through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
 source Location/Qualifiers
 1..639
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DQ1-clt-1-21-0-UI"
 /clone_1db="UI-R-DQ1"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DQ1
 library is a normalized Rat Cell Line R3327-5a library
 (nBSA) constructed in pT73D-Pac vector according to the
 procedure described by Bernaldo, Lemmon & Soares
 (Normalization and Subtraction: Two Approaches to
 Facilitate Gene Discovery. Genome Research 6: 791-806,
 1996). The oligonucleotide used to prime first strand
 synthesis contained the sequence tag GAGCTGATGATC between
 the Not I cloning site and dT18 stretch. The Rat Cell Line
 R3327-5A was provided by Mary Hendrix of the University of

Iowa.
 TAG LIB=UI-R-DQ1
 TAG TISSUE=rat cell line R3327-5a
 TAG SEQ=GGACTAGATC"

BASE COUNT 96 a 186 c 127 g 230 t

ORIGIN

Query Match 15.3%; Score 558.8; DB 14; Length 639;
 Best Local Similarity 93.3%; Pred. No. 5.3e-126;
 Matches 595; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 1303 GCCGCAAAACGGGAACTGAGAAAGCAGAGCACTTGAATGAGAAACGAGACCGAGAGAG 1362
 Db 639 GCAGCAAAACGGGAACTGAGAAAGCAGAGCACTTGAATGAGAAACGAGACCGAGAGAG 580
 QY 1363 GAATCTCTGAAATCAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1422
 Db 579 GAATCTCTGAAATCAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 520
 QY 1423 AAGATCTGAGAGTTGAGTTGAGAGCTGTAATGACAAAAAGCATTCAGCTAGAGAGAGAA 1482
 Db 519 AAGATCTGAGAGTTGAGTTGAGAGCTGTAATGACAAAAAGCATTCAGCTAGAGAGAGAA 461
 QY 1483 CTTGAGATATCAGAGTGTGAGCTGAGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1542
 Db 460 CTTGAGATATCAGAGTGTGAGCTGAGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
 QY 1543 TCTAGAGAGCTAGAGATGCTGTAATGACCACTTACAGAGAGAGAGAGAGAGAGAGAG 1602
 Db 400 TCTAGAGAGCTAGAGATGCTGTAATGACCACTTACAGAGAGAGAGAGAGAGAGAGAGAG 341
 QY 1603 CAATGCTTGAAGAGCTGTAATGACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1662
 Db 340 CAGATGCTTGAAGAGCTGTAATGACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 281
 QY 1663 CAGCAGAAACAGTTGTCATAGAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGAGAG 1722
 Db 280 CAGCAGAAACAGTTGTCATAGAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGAGAGAG 221
 QY 1723 GAGCTGCGCCCGGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1782
 Db 220 GAGCTGCGCCCGGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161
 QY 1783 AAGCTGAGAGAGATGATGATTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAG 1842
 Db 160 AAGCTGAGAGAGATGATGATTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 101
 QY 1843 AAACAGCAATCTCAG 1902
 Db 100 AAACAGCAATCTCAG 41
 QY 1903 GAGAGAAAGAGCTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
 Db 40 GAGAGAAAGAGCTGAGTTAG 3

RESULT 14
 BQ898718 1062 bp mRNA linear EST 16-AUG-2002
 LOCUS BQ898718
 DEFINITION AGENCOURT 8118759 lupski dorsal root ganglion Homo sapiens cDNA
 accession BQ898718
 version BQ898718.1 GI:22290732
 keywords EST.
 source human.
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1062)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.

Email: cgarbs-r@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM13562 row: n column: 16
 High quality sequence start: 116
 High quality sequence stop: 760.
 Location/Qualifiers

FEATURES

source

1..1062
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6180303"
 /clone_1ib="lupski_dorsal_root_ganglion"
 /sex="male"
 /tissue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6 (Life Technologies); Site_1:
 NotI; Site_2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TGACCCACGCGTCCG-3' and
 5'-GACTAGTTCTAGTCGCGCGCGCCCT(15)-3'. Size selected >
 1 kb for average insert length 1.7 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D (Baylor
 College of Medicine) and is available through Life
 Technologies."

BASE COUNT 338 a 250 c 261 g 209 t 4 others
 ORIGIN

Query Match 15.2%; Score 555; DB 14; Length 1062;

Best Local Similarity 90.1%; Pred. No. 5.5e-125;

Matches 594; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

OY 2984 TTCGCCGAGAGAGATTATTCATGTACACATACAGAGATTCTGAGCAAGAGATTAA 3043
 DB 113 TTCCGGGAGAGAGATTATTCATGTACACATACAGAGATTCTGAGCAAGAGATTAA 172
 OY 3044 CCTTCAGCAAGGGAGTGTATTGTGTTACCAAGAAAGTGTGTCGTCGACCGGAA 3103
 DB 173 CCTTCAGCAAGGGAGTGTATTGTGTTACCAAGAAAGTGTGTCGTCGACCGGAA 232
 OY 3104 CGGTGGCCGCAAGTCCGAGTCTTCCCTTCTACTATGTAGAGCTTAAAGATTCAAGG 3163
 DB 233 CAGTGGGCGCAAGGGCGGAGTCTTCCCTTCTACTATGTAGAGCTTAAAGATTCAAGG 292
 OY 3164 GCTCTGAACTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAAATTGCCAGGTTA 3223
 DB 293 GCTCTGAACTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAAATTGCCAGGTTA 352
 OY 3224 TTGCTTCTACGCTGTACTGTATCCGAAACAACCTGCTGCTGCGCAGCTGATT 3283
 DB 353 TTGCTTCTACGCTGTACTGTATCCGAAACAACCTGCTGCTGCGCAGCTGATT 412
 OY 3284 TGATCCGAAAAAAGAAACCAAGTGTGATGTTGGGAGAGAAATCGCAAGTCCGAGGAAAA 3343
 DB 413 TGATCCGAAAAAAGAAACCAAGTGTGATGTTGGGAGAGAAATCGCAAGTCCGAGGAAAA 472
 OY 3344 AGGCGAGATAGGGGTGTTCCAGCAAAATTATGTCAAACTTCAACCCCGGAAACAAGA 3403
 DB 473 AGGCGAGATAGGGGTGTTCCAGCAAAATTATGTCAAACTTCAACCCCGGAAACAAGA 532
 OY 3404 AAATCAACCCCACTGAGTACCCAGACCGCAGTGCAGCAGCAGTGTGCCAGTGTATG 3463
 DB 533 AAATCACTCCAAAGAGCAACCTAAGTCAACACATTAGGGGAGTGTGCCAGTGTATG 592
 OY 3464 GATGTATGATTAACCGCCGCAAGAGATGACAACTAGCTTCAAGCAAAAGGCAGATGA 3523
 DB 593 GATGTATGATTAACCGCCGCAAGAGATGAGTGTGCTTCAACAAAGGCAGATGA 652

OY 3524 TCAACGTCCTCAACAGAGAGACCCGACTGTGTGAAAAAGAGAAATCAGTGGCAAGTTG 3583
 DB 653 TCAACGTCCTCAACAGAGAGACCCGACTGTGTGAAAAAGAGAAATCAGTGGCAAGTTG 712
 OY 3584 GGCCTTCCCATCAATTAATGTAAGCTGACACAGCATGAGACCCGACAGCAATGA 3642
 DB 713 GGCCTTCCCATCAATTAATGTAAGCTGACACAGCATGAGACCCGACAGCAATGA 771

RESULT 15

LOCUS

BE373101 724 bp mRNA linear EST 21-JUL-2000

DEFINITION 601224527F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582726 5',

ACCESSION BE373101

VERSION BE373101.1 GI:9318464

KEYWORDS

SOURCE

ORGANISM

TITLE

AUTHORS

REFERENCE

JOURNAL

COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgarbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM6738 row: f column: 07
 High quality sequence stop: 571.
 Location/Qualifiers

FEATURES

source

1..724
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3582726"
 /clone_1ib="NCI CGAP Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT

181 a 185 c 245 g 113 t

Query Match

14.8%; Score 537.8; DB 10; Length 724;

Best Local Similarity 97.9%; Pred. No. 7.8e-121;

Matches 608; Conservative 0; Mismatches 7; Indels 6; Gaps 6;

OY 695 ACGCAAAACTATAGTGTGACACTTAAACAGTCCCGCAGCAAGAACTATCTCATGCAAT 754
 DB 1 ACGCAAAACTATAGTGTGACACTTAAACAGTCCCGCAGCAAGAACTATCTCATGCAAT 60
 OY 755 CAAGTTAACCCAGAGCTCAGTGTGCTCAATATAGGAATCTTTGACATGATGCAAGATG 814
 DB 61 CAAGTTAACCCAGAGCTCAGTGTGCTCAATATAGGAATCTTTGACATGATGCAAGATG 119
 OY 815 GAAATCACTGAGAGAAATTTATCTAGCTATGCACTTAATGATGTCATGTCG 874
 DB 120 GAAATCACTGAGAGAAATTTATCTAGCTATGCACTTAATGATGTCATGTCG 179
 OY 875 GTACAGCACTGCGCGCTGCTGCTCCAGAAATGATCCCTCTTCTTCAAGAGATT 934
 DB 180 GTACAGCACTGCGCGCTGCTGCTCCAGAAATGATCCCTCTTCTTCAAGAGATT 239

QY 935 GCTCCGACAGTGGATGTCCGTACATAAGCTCTTCTTCTGTGATCAGAGCTGCTGAG 994
 Db 240 GCTCCGACAGTGGATGTCCGTACATAAGCTCTTCTTCTGTGATCAGAGCTGCTGAG 299
 QY 995 AGCGTGTCAAGATGAGCAGCAGCCAGAGAACTTGCTGTGACATTTGAAGATA 1054
 Db 300 AGCGTGTCAAGATGAGCAGCAGCCAGAGAACTTGCTGTGACATTTGAAGATA 357
 QY 1055 AGAAGCGGAGAACTTGTGAGGAGCAGTGTGAGCTGAGAAAGCCCGCAAGCGCTCT 1114
 Db 358 AGAAGCGGAGAACTTGTGAGGAGCAGTGTGAGCTGAGAAAGCCCGCAAGCGCTCT 416
 QY 1115 TGGAGCAGCAGCGCAAGAGCAGGAGCGGTGCTCAGCTGAGCGCGCCAGCAGGAGA 1174
 Db 417 TGGAGCAGCAGCGCAAGAGCAGGAGCGGTGCTCAGCTGAGCGCGCCAGCAGGAGA 476
 QY 1175 GGAAGAGCGGAGCGCCAGGAGCAGAG--GCCAAGCGGCACTGTGAGCTGTGAGAAACAG 1233
 Db 477 GGAAGAGCGGAGCGCCAGGAGCAGAGCGCAATGCGGCACTGTGAGCTGTGAGAAACAG 536
 QY 1234 CTGAGAGCAGCGGAGCTGTGAGCGGAGCAGGAGAGAGAGAGAGAGAGAGATGAG 1293
 Db 537 CTGAGAGCAGCGGAGCTGTGAGCGGAGCAGGAGAGAGAGAGAGAGAGAGAG--GATGAG 595
 QY 1294 AGCGCGAGCGCCAGAAACGG 1314
 Db 596 ACGCGAGCGCCAGAAAGGGG 616

Search completed: March 14, 2003, 11:05:44
 Job time : 4408.26 secs

PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 PS Claim 14; Page 43; 99pp; English.
 CC The present sequence represents mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Esp1s protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its antagonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp1s complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX
 SQ Sequence 1214 AA:
 Query Match 99.8%; Score 6258.5; DB 21; Length 1214;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1213; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 600 DFNNOULKELEIHSKOOLOKORSTFAARLUKOKROEKSLLELEKOKEDAOARVQERPKOM 659
 Db 601 DFNNOULKELEIHSKOOLOKORSTFAARLUKOKROEKSLLELEKOKEDAOARVQERPKOM 660
 QY 660 LEHVQOEOPRRPKPHEDRLKREDSVRKKEAEERAKPENODKOSRLFHPHOEPAKLATQ 719
 Db 661 LEHVQOEOPRRPKPHEDRLKREDSVRKKEAEERAKPENODKOSRLFHPHOEPAKLATQ 720
 QY 720 APWSTTEGPIITIAQOSVKVYVRALYPRESRSHDITIQPGDIVWVDSQGEPEWMLG 779
 Db 721 APWSTTEGPIITIAQOSVKVYVRALYPRESRSHDITIQPGDIVWVDSQGEPEWMLG 780
 QY 780 GELGKGTMPFANYAEKIPENEVTPPAKPYTDLTSAAPAKLARETPAPLPVTSSEBSTT 839
 Db 781 GELGKGTMPFANYAEKIPENEVTPPAKPYTDLTSAAPAKLARETPAPLPVTSSEBSTT 840
 QY 840 PNNMADFSTWSSSNEKETDNDWTAQPSLTVPASAGLROKRSATPATATGSSPSPV 899
 Db 841 PNNMADFSTWSSSNEKETDNDWTAQPSLTVPASAGLROKRSATPATATGSSPSPV 900
 QY 900 LGQGEKVGLQAQALYPRAKDNHLPNKSVDITVLEQDDMMWFGVQOKGMFPKSYV 959
 Db 901 LGQGEKVGLQAQALYPRAKDNHLPNKSVDITVLEQDDMMWFGVQOKGMFPKSYV 960
 QY 960 KLISGPVKSTSIDTGPSPASLKRVASPAKPAIPGEETIAMTYESESSEOGDLTPQOG 1019
 Db 961 KLISGPVKSTSIDTGPSPASLKRVASPAKPAIPGEETIAMTYESESSEOGDLTPQOG 1020
 QY 1020 DVIIVTKKDGDMWTGTVGDKSGVFPSSNVYRLKDSGSGTGAKTGSJCKKPEIAOVIASYA 1079
 Db 1021 DVIIVTKKDGDMWTGTVGDKSGVFPSSNVYRLKDSGSGTGAKTGSJCKKPEIAOVIASYA 1080
 QY 1080 ATGPEQLTLAPGOLILIRKKNPGGWMEGELQARGKKQIGFPPANYVKLSPGSKITPT 1139
 Db 1081 ATGPEQLTLAPGOLILIRKKNPGGWMEGELQARGKKQIGFPPANYVKLSPGSKITPT 1140
 QY 1140 ELKPTAVQPAVCQVIGMYDYTAQNDDELAFSGQIINVLNKEDPMMWKGVSQVGLFPS 1199
 Db 1141 ELKPTAVQPAVCQVIGMYDYTAQNDDELAFSGQIINVLNKEDPMMWKGVSQVGLFPS 1200
 QY 1200 NYVKLTITMDPSQQ 1213
 Db 1201 NYVKLTITMDPSQQ 1214

RESULT 2
 ID AAY57449 standard; Protein: 1715 AA.
 XX
 AC AAY57449;
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Ese1 protein sequence.
 XX
 KW Mouse; murine; Ese1, Ese2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 XX antiviral.
 OS Mus sp.
 XX
 PN W09955728-A2.
 XX
 PD 04-NOV-1999.
 XX
 XX 27-APR-1999; 99WO-CA00375.
 PF
 XX 27-APR-1998; 98CA-2230201.
 PR
 XX 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;
 PI
 DR WPI: 2000-052802/04.
 DR N-PDSB; AA339024, AA339025.
 XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 PS Claim 14; Page 62-63; 99pp; English.
 XX
 CC The present invention specifically describes mammalian Ese1 and 2
 CC proteins (1) and their splice variants (Ese = Eh-domain and SH3-domain
 CC regulator of endocytosis). (1) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Eps15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (1) (or its antagonists,
 CC mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab);
 CC sequences antisense to the (1) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with underexpression of endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (1) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (1) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese1 protein sequence.
 CC
 XX

Sequence 1715 AA:

Query Match 99.8%; Score 6258.5; DB 21; Length 1715;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1213; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAQPPPPGSLVVAITVERAKHQOFLSLKPIAGFITGDARNFPOSGLPQVLAQ 60
 Db 1 MAQPPPPGSLVVAITVERAKHQOFLSLKPIAGFITGDARNFPOSGLPQVLAQ 60
 QY 61 IVALADNNNGRMDQVEFSIAMKLILKLOGYOLPSTLPVMKQOFAISSAPAFIGGI 120
 Db 61 IVALADNNNGRMDQVEFSIAMKLILKLOGYOLPSTLPVMKQOFAISSAPAFIGGI 120
 QY 121 ASMPPLTAAVAVPMGSIPIVVGMSPLVSSVPPAAVPLANGAPVPIQPLPAFAHPAATWP 180
 Db 121 ASMPPLTAAVAVPMGSIPIVVGMSPLVSSVPPAAVPLANGAPVPIQPLPAFAHPAATWP 180
 QY 181 KSSSFSRSGSGSGLNTKLOKAGSFVVASAPPAEMAVPQSSRLKRYQLRNSHDKTMSGHL 240
 Db 181 KSSSFSRSGSGSGLNTKLOKAGSFVVASAPPAEMAVPQSSRLKRYQLRNSHDKTMSGHL 240
 QY 241 TGPQARTILMOSSLPOKQALASINWLSIDIDDGKLTAEFLAHMLLDVAMSQGLPVL 300
 Db 241 TGPQARTILMOSSLPOKQALASINWLSIDIDDGKLTAEFLAHMLLDVAMSQGLPVL 300
 QY 301 PEYIPPSFRFRVSGSGSVISSSVDOQLPEPSSSEDEQPEKKLPVTFEKKRENFERG 360
 Db 301 PEYIPPSFRFRVSGSGSVISSSVDOQLPEPSSSEDEQPEKKLPVTFEKKRENFERG 360
 QY 361 SVELEKRRQALLQORKEQERLAQLERADEKREKREKREKREKREKREKREKREKREK 420
 Db 361 SVELEKRRQALLQORKEQERLAQLERADEKREKREKREKREKREKREKREKREKREK 420
 QY 421 QREERKEKEIEERPAKRELEKROLEWERNRROELLNQNKEOEGTVLKARKTLLEFE 480
 Db 421 QREERKEKEIEERPAKRELEKROLEWERNRROELLNQNKEOEGTVLKARKTLLEFE 480
 QY 481 LEALNDKQHOLEGKLODIRCLATQROEISTYNSRELRIAEITTHLQOQLQESQOMLGR 540

Db 481 LEALNDKQHOLEGKLODIRCLATQROEISTYNSRELRIAEITTHLQOQLQESQOMLGR 540
 QY 541 IPEKQILSDQKQVQNSLHRDSILTLKRALLEAKELARQOLREQDVEVERETRSKLOEID 600
 Db 541 IPEKQILSDQKQVQNSLHRDSILTLKRALLEAKELARQOLREQDVEVERETRSKLOEID 600
 QY 601 VFNNQKELREIHSKQOLQKORSLEAARLKQKEQERKSELEKQEKEDAORRVERDKQWL 660
 Db 601 VFNNQKELREIHSKQOLQKORSLEAARLKQKEQERKSELEKQEKEDAORRVERDKQWL 660
 QY 661 EHVQOEEDPRPKRHEEDRLKREDSVRKKEAEBAKEMODKOSRLPHPOEPAKLATQA 720
 Db 661 EHVQOEEDPRPKRHEEDRLKREDSVRKKEAEBAKEMODKOSRLPHPOEPAKLATQA 720
 QY 721 PMSTTEKPLTISAQESVKKVYVYALYPFESRSHDITTIQPGDIVMDESQTEGPGWLG 780
 Db 721 PMSTTEKPLTISAQESVKKVYVYALYPFESRSHDITTIQPGDIVMDESQTEGPGWLG 780
 QY 781 ELKGTGWFPAVNAEKIPENEVPTPAKPVTDLTSAAPAKLARETPAPLPVTSSESTTP 840
 Db 781 ELKGTGWFPAVNAEKIPENEVPTPAKPVTDLTSAAPAKLARETPAPLPVTSSESTTP 840
 QY 841 NNWADFSTWSSSNKREPTDNMDTMAAOPSLTVPSAGOLRQSAFTPATATGSSPSVL 900
 Db 841 NNWADFSTWSSSNKREPTDNMDTMAAOPSLTVPSAGOLRQSAFTPATATGSSPSVL 900
 QY 901 GQGEVEGLQALYPPRAKKNHNFNKSVDITVLEQODMMWFGVEQOGKMPKSYVK 960
 Db 901 GQGEVEGLQALYPPRAKKNHNFNKSVDITVLEQODMMWFGVEQOGKMPKSYVK 960
 QY 961 LISGVRKSTSIDTGPTEPSASLKRVASPAKPAIPEEPIAMTYESSQGLTFQOGD 1020
 Db 961 LISGVRKSTSIDTGPTEPSASLKRVASPAKPAIPEEPIAMTYESSQGLTFQOGD 1020
 QY 1021 VIYVT-KKDGWMTGTGDKSGVFPSSVYVRLKDESGGTAGKSLKKREIAQVIAVSYA 1079
 Db 1021 VIYVT-KKDGWMTGTGDKSGVFPSSVYVRLKDESGGTAGKSLKKREIAQVIAVSYA 1079
 QY 1080 ATGPQTLTAPGOLILRKKNPGGMEGELQARCKKQIGFPAVNYKLSPGSKITPT 1139
 Db 1080 ATGPQTLTAPGOLILRKKNPGGMEGELQARCKKQIGFPAVNYKLSPGSKITPT 1139
 QY 1141 ELPKTAQVPAVCQYIGWYDTAQNDELAFSGQIINVLNKEDPDMKGEVSGVGLFPS 1200
 Db 1141 ELPKTAQVPAVCQYIGWYDTAQNDELAFSGQIINVLNKEDPDMKGEVSGVGLFPS 1200
 QY 1200 NYVKLTMDPSQ 1213
 Db 1201 NYVKLTMDPSQ 1214

RESULT 3

AA332156
 ID AA332156 standard; Protein; 1215 AA.

AA332156;

DT 01-FEB-2000 (first entry)

XX Human SH3D1A protein.

XX SH3D1A; human; Down's syndrome; leukemia; cancer;
 XX megakaryocytic abnormality; myeloproliferative disorder;
 XX platelet disorder; neural disorder; thrombocytopenia;
 XX haematopoietic disorder; cognitive dysfunction; microcephaly;
 XX lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.

OS Homo sapiens.

XX

XX

PN W09953062-A2.

XX 21-OCT-1999.

PD

XX PF 16-APR-1999; 99WO-US08371.
 XX PR 16-APR-1998; 98US-0082007.
 XX PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX PI Korenberg JR, Chen X;
 XX DR WPI: 1999-633829/54.
 XX DR N-PSDB; AA234572.
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 XX PS
 PS Claim 14; Fig 11; 99pp: English.
 CC This sequence represents the protein encoded by the human SH3D1A
 CC cDNA clone 11 (see AA234572). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AA234572-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided: for suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.
 CC XX
 SO Sequence 1215 AA;
 Query Match 93.5%; Score 5860; DB 20; Length 1215;
 Best Local Similarity 92.9%; Pred. No. 0;
 Matches 1129; Conservative 39; Mismatches 45; Indels 2; Gaps 2;

Db 361 GNLLEKRRQALLLEQRRKEQRLAQLERAEQERKERERQERKROLELEKORELE 420
 QY RQREERERKEITERRAARELEERQOLEMERENROELLNQNKEQEGVYVAKARKTLEF 479
 Db 421 RQREERERKEITERRAARELEERQOLEMERENROELLNQNKEQOEDI VVAKARKTLEF 480
 QY 480 ELEALNDKKHOLEGLQDIRGLATORQIESTYKSEBELIAETIHLQOQLOESQOMLGR 539
 Db 481 ELEALNDKKHOLEGLQDIRGLATORQIESTYKSEBELIAETIHLQOQLOESQOMLGR 540
 QY 540 LIPEKQIISDQIKOYQONS LHRDS LITLTKALAEKELARQOLREQLDVEKETSKIOEI 599
 Db 541 LIPEKQIINDQIKOYQONS LHRDS LITLTKALAEKELARQOLREQLDVEKETSKIOEI 600
 QY 600 DVENNQLKELEIHSKQOLQORSLEAARLQKQOEKSELELEKQEDACORVOERDKOM 659
 Db 601 DIFNNQLKELEIHSKQOLQORSLEAARLQKQOEKSELELEKQEDACORVOERDKOM 660
 QY 660 LEHVOQE-EQPRPRKPHBEDRLKEDSVYRKKAEERAKPEMOKQSLFPHQEPKALAT 718
 Db 661 LEHVOQEDHQRPRLKHEBEKLEESYKKGGEKQOEADKLGRLFHQHPAPKAV 720
 QY 719 QAPWSTERKPLTISAQSVVYVYRALYPRSSHDEITIQPDIVMWDSQTEGEGWL 778
 Db 721 QAPWSTERKPLTISAQSVVYVYRALYPRSSHDEITIQPDIVMWDSQTEGEGWL 780
 QY 779 GGELGKGTGWPFPANYAEKI PENEVPTPAKPYTDLTSAAPKALARETAPLPVTSSEPT 838
 Db 781 GGELGKGTGWPFPANYAEKI PENEVPTPAKPYTDLTSAAPKALARETAPLPVTSSEPT 840
 QY 839 TPNNMADSSSTWPPSSSNKPEETDMDWTAAOPSLTVBSAGOLRORSAPTPATATGSSPSP 898
 Db 841 TPNNMADSSSTWPPSSSNKPEETDMDWTAAOPSLTVBSAGOLRORSAPTPATATGSSPSP 900
 QY 899 VLGQGEKTEGLOAQALYWRRAKKNHNLFNKSDVITVLEQDDMMWFGVQOGKMPFKSY 958
 Db 901 VLGQGEKTEGLOAQALYWRRAKKNHNLFNKSDVITVLEQDDMMWFGVQOGKMPFKSY 960
 QY 959 VKLISGPVRKSTSIDTGTBESPASIKRYASPAKPAIDGEEFIAMYYESSEQDLPFOQ 1018
 Db 961 VKLISGPVRKSTSIDTGTBESPASIKRYASPAKPAIDGEEFIAMYYESSEQDLPFOQ 1020
 QY 1019 GDIIVTETKKDGMWGTGTDGSKGVFPSPSYVRLKDSGSGTGKSGIKKPEIKQVASY 1078
 Db 1021 GDIIVTETKKDGMWGTGTDGSKGVFPSPSYVRLKDSGSGTGKSGIKKPEIKQVASY 1080
 QY 1079 AATGPEQLTLAPGQILIRKKNPGGMEGELQARQKROIGWFPANYVKLLSPGTSKITP 1138
 Db 1081 AATGPEQLTLAPGQILIRKKNPGGMEGELQARQKROIGWFPANYVKLLSPGTSKITP 1140
 QY 1139 TELPKTAQPAVCQVIGYVDYTAQNDDELAFSKQIINVLNKEPDMWKGKGVSGVGLFP 1198
 Db 1141 TELPKTAQPAVCQVIGYVDYTAQNDDELAFSKQIINVLNKEPDMWKGKGVSGVGLFP 1200
 QY 1199 SNVYKLTITDMDPSQ 1213
 Db 1201 SNVYKLTITDMDPSQ 1215
 RESULT 4
 AA32155
 ID AA32155 standard; Protein, 1220 AA.
 XX AA32155;
 AC
 XX 01-FEB-2000 (first entry)
 DT
 XX Human SH3D1A protein.
 DE
 XX SH3D1A: human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW

XX 01-FEB-2000 (first entry)
 DT Human SH3D1A protein.
 XX
 XX SH3D1A; human; Down's syndrome; leukaemia; cancer;
 KM megakaryocytic abnormality; myeloproliferative disorder;
 KM platelet disorder; neural disorder; thrombocytopenia;
 KM haematopoietic disorder; cognitive dysfunction; microcephaly;
 KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 15..102
 FT Domain /note= "EH1 domain"
 FT Domain 215..310
 FT Domain /note= "EH2 domain"
 FT Domain 740..800
 FT Domain /note= "SH3-1 domain"
 FT Domain 908..966
 FT Domain /note= "SH3-2 domain"
 FT Domain 999..1062
 FT Domain /note= "SH3-3 domain"
 FT Domain 1080..1138
 FT Domain /note= "SH3-4 domain"
 XX
 PN W0995062-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 16-APR-1999; 99WO-US08371.
 XX
 PR 16-APR-1998; 98US-0082007.
 XX
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 PI Korenberg JR, Chen X;
 XX
 DR WPI: 1999-633829/54.
 DR N-PSDB; AA234570.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 XX
 PS Claim 14; Fig 5; 99PB; English.
 XX
 CC This sequence represents the protein encoded by the novel human
 CC SH3D1A gene (see AA234570). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukaemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AA23154-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC chromosomal disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided: for suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality; myeloproliferative disorder;
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.
 XX
 SQ Sequence 1144 AA;

Query Match 86.8%; Score 5442.5; DB 20; Length 1144;
 Best Local Similarity 87.2%; Pred. No. 0;
 Matches 1060; Conservative 37; Mismatches 45; Indels 73; Gaps 3;
 QY 1 MAQPTPPGGSLDVAIIVVEERAKHDQFLSKPLAGITTDQKRNFPQSGLPQPLVLAQ 60
 DB 1 MAQPTPPGGSLDVAIIVVEERAKHDQFHSKPLSGITTDQKRNFPQSGLPQPLVLAQ 60
 QY 61 IVALADNNNDGMDQVFSIAKLIKLYGQLESTLPVYKQOQVAISSAPAFGIGI 120
 DB 61 IVALADNNNDGMDQVFSIAKLIKLYGQLESTLPVYKQOQVAISSAPAFGIGI 120
 QY 121 ASMPPLTAAVAPVPMGSIYVVGKSPVLSVPPAAVPLANAPVYIOPLPAFAHPATLP 180
 DB 121 ASMPPLTAAVAPVPMGSIYVVGKSPVLSVPPAAVPLANAPVYIOPLPAFAHPATLP 180
 QY 181 KSSFSRSGPQSLNTKLOKAQSPVAGAPPAEWAAPQSSRLKYROLFNSHDXTMSGHL 240
 DB 181 KSSFSRSGPQSLNTKLOKAQSPVAGAPPAEWAAPQSSRLKYROLFNSHDXTMSGHL 240
 QY 241 TGPQARTILMOGSLPQALASTIWNLSIDIDODGKLTAEFFILAMHLIDVAMSGOPLPVLP 300
 DB 241 TGPQARTILMOGSLPQALASTIWNLSIDIDODGKLTAEFFILAMHLIDVAMSGOPLPVLP 300
 QY 301 PEYIPSPFRVAVSGSGMSVSSVSDORLPPEPSSDEDOQP-EKKLPVTFEDKKRENER 359
 DB 301 PEYIPSPFRVAVSGSGMSVSSVSDORLPPEPVLDEDOQLEKKLPVTFEDKKRENER 360
 QY 360 GSVELEKRRQALLLEQORKEOERLAOLEBAOEERKEOERKEOERKEOERLEKOLEKORELE 419
 DB 361 GNLELEKRRQALLLEQORKEOERLAOLEBAOEERKEOERKEOERKEOERLEKOLEKORELE 420
 QY 420 RQREERREKETERRAARELEERQOLEMERNRROELINQNKQOEGVYLKARKKTLF 479
 DB 421 RQREERREKETERRAARELEERQOLEMERNRROELINQNKQOEDVLKARKKTLF 480
 QY 480 ELEALNDKKHGLEGLQDIRGLATOROEISTNKSRELTAETHTHLOOQLOESQOMLGR 539
 DB 481 ELEALNDKKHGLEGLQDIRGLATOROEISTNKSRELTAETHTHLOOQLOESQOMLGR 540
 QY 540 LIPEKQIISDQKQVQNSLHRDLSLTLKALBAKELARQOLREOLDEVEETSKLOEI 599
 DB 541 LIPEKQIISDQKQVQNSLHRDLSLTLKALBAKELARQOLREOLDEVEETSKLOEI 600
 QY 600 DVFNNOLEKREIHSKQOLQORSLEAARLKQKEOERKLELEKQKEPAORRQVBRDKQ 659
 DB 601 DVFNNOLEKREIHSKQOLQORSLEAARLKQKEOERKLELEKQKEPAORRQVBRDKQ 660
 QY 660 LEHVQOE-EQPRPRKPHBEDRLKREDSYRKKEAERAKPEMOKQSRLEFHHQEPKALAT 718
 DB 661 LEHVQOEDHQPRKLEHEEKLKREESYKKGDEKQOEADKLGRLFHHQEPKAPAV 720
 QY 719 QAPWSTTEKGPLTIAQSSVYVYVYRALYPPESRSHDITTOPIQDIWVNDSSQGEPCWL 778
 DB 721 QAPWSTTEKGPLTIAQSSVYVYVYRALYPPESRSHDITTOPIQDIWVNDSSQGEPCWL 780
 QY 779 GSELKQKTCMPFANYAEKIPENEVPTPAKPVTDLSAPAPKLAURETPAPPVTSSEST 838
 DB 781 GSELKQKTCMPFANYAEKIPENEVPTPAKPVTDLSAPAPKLAURETPAPPVTSSEST 840
 QY 839 TNNNMADESSWPSSNEKPEETDNWDTAAOPLTVPSAGOLRQSAFTPATAGSSSP 898
 DB 841 TNNNMADESSWPSTNKEPETDNWDTAAOPLTVPSAGOLRQSAFTPATAGSSSP 900
 QY 899 VVGQGEKVEGLOALVYPRAKDNHLNFKNSDVTTLVEQDDMMVFGVQOKGMFPFSY 958
 DB 901 VVGQGEKVEGLOALVYPRAKDNHLNFKNSDVTTLVEQDDMMVFGVQOKGMFPFSY 960
 QY 959 VKLISGPVYKSTSIDTGTESPASLKRVAAPAKPAIPGEEFIAMTYTESSEQDGLTFQO 1018
 DB 961 VKLISGPVYKSTSIDTGTESPASLKRVAAPAKPAIPGEEFIAMTYTESSEQDGLTFQO 1000

QY 1019 GDVIVTKKDDMMGTGVDKSGVPSNYRLKDSGSGTACKTSGLGKKPEIAQVIASY 1078
Db 1001 -----ETIAQVIASY 1009
QY 1079 AATGPOLTLAPGOLLIRKKNPGWMEGLQARKKROI GMPANVYKLISGTSKITP 1138
Db 1010 TATGPOLTLAPGOLLIRKKNPGWMEGLQARKKROI GMPANVYKLISGTSKITP 1069
QY 1139 TELPKTAVGPVACQVI GMYDYTAQNDELAFSKQI INVLNKEDPDMWKEVSGQVGLFP 1198
Db 1070 TEPKSTALAAVQVIGMYDYTAQNDELAFNKGQI INVLNKEDPDMWKEVSGQVGLFP 1129
QY 1199 SNVYKLTMDPSQ 1213
Db 1130 SNVYKLTMDPSQ 1144
RESULT 6
AAU87168
ID AAU87168 standard; Protein; 1031 AA.
XX
AC AAU87168;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #78.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240950.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

XX
AC AAM4519;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 197.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; anti-allergic; hepatocytic; antidiabetic;
KW antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;
KW antiparasitic; cardiast; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
XX
OS Homo sapiens.
XX
PN WO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228824.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249289.
PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-488781/53.
 DR N-PSDB; AA163825.
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -
 XX
 XX Claim 11, SEQ ID NO 197; 664pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA163803-AA164012) and
 CC the encoded proteins (AA443497-AA443660) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 1035 AA:
 Query Match 68.2%; Score 4277; DB 22; Length 1035;
 Best Local Similarity 92.3%; Pred. No. 3,4e-266;
 Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;

Db 308 PVLPEYTPPSFRRVRSRSGSVSVSSVSDQRLPEEPVLEDEQQLEKKLPVTFEDKKRE 367
 QY 356 NFERGSELEKRRQALLLEQKREORLAQLERAEQEKEREREOEAKRLLELEKOLEKQ 415
 Db 368 NFERGNLELEKRRQALLLEQKREORLAQLERAEQEKEREREOEAKRLLELEKOLEKQ 427
 QY 416 RELEROREERERKEIERERAKRELEROROLEMRNRROELNORNEOGSTVVLKARRK 475
 Db 428 RELEROREERERKEIERERAKRELEROROLEMRNRROELNORNEOGSTVVLKARRK 487
 QY 476 TLEFELALNDKQQLBGLQDIFRCRLATOROEISTNKSRELRIAETHLOOQLOESQ 535
 Db 488 TLEFELALNDKQQLBGLQDIFRCRLATOROEISTNKSRELRIAETHLOOQLOESQ 547
 QY 536 MLEGRILPEKQILNDQLOKVOONSJHRDSILTLKALAKELARQQLPEQIDVEERETR 595
 Db 548 MLEGRILPEKQILNDQLOKVOONSJHRDSILTLKALAKELARQQLPEQIDVEERETR 607
 QY 596 LOEIDVFNNOQLEKEIREIHSKQOLKORSLEAARLKOEKERSLELEKOKEDAQRRVOER 655
 Db 608 LOEIDVFNNOQLEKEIREIHSKQOLKORSLEAARLKOEKERSLELEKOKEDAQRRVOER 667
 QY 656 DKOWLEHVQOE-EOPRPKPHEEDRLREDSVKKKEABERAKPMDOKOSRLLFPHQEP 714
 Db 668 DKOWLEHVQOEDEHQRPKXLEHEEKLKREESVKKKDEEKOKOAPQDKRLFPHQEPA 727
 QY 715 KLATQAPWSTTEKGPLTISAQESVKKVYVYRALYPFESRSHDEITOGDITVMDVESQTGE 774
 Db 728 KPVAQAPSTBEKGPLTISAQENKVVYRALYPFESRSHDEITOGDITVMDVESQTGE 787
 QY 775 PGMGLGELKGTGWFPANVYAEKIPENEVPTPAKVDTLTSAPAKLALRETPAPLPVTSS 834
 Db 788 PGMGLGELKGTGWFPANVYAEKIPENEVPAVKPVTSTSAAPAKLALRETPAPLPVTSS 847
 QY 835 EESTTPNNWAPSSSWPSSSNEKEPTDMDWAAPSLTVPSAGQLRORSAFTPATATGS 894
 Db 848 EESTTPNNWAPSSSWPSTVNEKEPTDMDWAAPSLTVPSAGQLRORSAFTPATATGS 907
 QY 895 SPSPVLAGQ 902
 Db 908 SPSPVLAGQ 915
 RESULT 8
 ID ABG20575 standard; Protein; 932 AA.
 XX
 AC ABG20575;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #20566.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS84762.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20, SEQ ID No 50934, 103pp: English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_sequences.
XX
SQ Sequence 932 AA;
Query Match 49.4%; Score 3097.5; DB 22; Length 932;
Best Local Similarity 80.0%; Pred. No. 2e-190;
Matches 603; Conservative 26; Mismatches 26; Indels 99; Gaps 3;
QY 461 NKEQEGVVLKARKKTEFELEALNDKKHOLEGKLDIRCLATROEISTNKSRELRI 520
DB 2 NKEQEDIVVLKAKKKTLEFELEALNDKKHOLEGKLDIRCLATROEISTNKSRELRI 61
QY 521 AEITHLQOOLQESQOMGRLLPEKQILSDOLKVOQNSLHRDILLTKRLALEKELARQ 580
DB 62 AEITHLQOOLQESQOMGRLLPEKQILSDOLKVOQNSLHRDILLTKRLALEKELARQ 121
QY 581 LREOLDEVERETRSKLOEIDVFNQNLKELREIHSKQOLKQKSLAARLQKQEKRSLE 640
DB 122 LRDLDEVEKETRSKLOEIDVFNQNLKELREIHSKQOLKQKSLAARLQKQEKRSLE 181
QY 641 LERKEDARRRVQERDQYMLEHVOQ-EQPPRRKPHEDRLKREDSVYRKKEAEERAKPEM 699
DB 182 LERKEDARRRVQERDQYMLEHVOQ-EQPPRRKPHEDRLKREDSVYRKKEAEERAKPEM 241
QY 700 QDKQSRLEFHPRQEPKATQAPWSTTEKGRPLTSAQESVVVVYVYRAIYPPRESSHDIT 759
DB 242 QDKQSRLEFHPRQEPKATQAPWSTTEKGRPLTSAQESVVVVYVYRAIYPPRESSHDIT 301
QY 760 QPGDIVWVDESQTEGPMGLGELKGTGMPFANYAEKI PENEVPTPAKPYTDTSAAPAK 819
DB 302 QPGDIVWVDESQTEGPMGLGELKGTGMPFANYAEKI PENEVPTPAKPYTDTSAAPAK 361
QY 820 LALRETPAPLPVTSSEPTTPNNMADFSSTWSSSNKPEPTDMDWTAAQPSLTVPASQ 879
DB 362 LALRETPAPLPVTSSEPTTPNNMADFSSTWSSSNKPEPTDMDWTAAQPSLTVPASQ 421
QY 880 LROSAFTPATGSSPSPLVGGEGEGLQAOALYPMRAKKNHINFNKSDVITVLEEQ 939
DB 422 LROSAFTPATGSSPSPLVGGEGEGLQAOALYPMRAKKNHINFNKSDVITVLEEQ 481
QY 940 DMWVFGEVQKGMFKPSYVVLISGVPYRKSTSIDGTESBASLKRVASPAKPAIFGEE 999
DB 482 DMWVFGEVQ-----E 491
QY 1000 FIAMTYESESQGLTFQGGDIVVITKKDGMWTGTVGDKSGVFPSPNYVRLKDSGSGTA 1059

DB 492 FIAMTYESESQGLTFQGGDIVVITKKDGMWTGTVDKAGVFPSPNYVRLKDSGSGTA 551
QY 1060 GKTGSLGKKPELQVYASVYATGPQTLAPQQLILIRKKNCGWMEGLQARKKRIQIG 1119
DB 552 GKTGSLGKKPELQVYASVYATGPQTLAPQQLILIRKKNCGWMEGLQARKKRIQIG 593
QY 1120 WEPANYVXLLSPGSKITPTTELPRKTAQVAVCOVIGWYDTPAONDELAFSKQIINVLN 1179
DB 594 -----VCQVIGWYDTPAONDELAFSKQIINVLN 623
QY 1180 KEDPDMWKGEGVSGVGLPSPNYVXLLTMDPSQ 1213
DB 624 KEDPDMWKGEGVSGVGLPSPNYVXLLTMDPSQ 657
RESULT 9
AA017881
ID AA017881 standard; Protein, 1681 AA.
XX
AC AA017881;
XX
DT 22-AUG-2002 (first entry)
XX
DE Allergic disease examination method related human protein.
XX
KW Allergic disease; allergy; antiallergic; intersectin 2; eosinophil;
XX atopic dermatitis; human.
XX Homo sapiens.
XX OS
XX PN WO200233122-A1.
XX PD 25-APR-2002.
XX PF 11-OCT-2001; 2001WO-JP08937.
XX PR 13-OCT-2000; 2000JP-0314093.
XX PA (GENO-) GENOX RES INC.
XX PA (NIGE-) JAPAN GEN NAT CHILDREN'S HOSPITAL.
XX PA (EISA) EISAI CO LTD.
XX Sugiata Y, Hashida R, Ogawa K, Obayashi M, Nagasu T, Saito H;
XX Takahashi E;
XX DR MPI: 2002-372313/40.
XX DR N-PSDB; AAL47247.
XX
PT Method for examining allergic diseases by differential display of
PT intersectin 2 gene showing different expression particularly
PT significant increase in eosinophils in patients
XX
PS Disclosure: Page 78-86, 90pp; Japanese.
XX
CC The present invention relates to a method for examining allergic diseases
CC with intersectin 2 gene or a gene with equivalent function of intersectin
CC 2 as an indicator gene, which comprises determining the expression level
CC of the gene in the eosinophils in a patient, and comparing the expression
CC level with that in the eosinophils of a healthy individual. The method is
CC also applicable in screening candidate compounds for remedies. The
CC present sequence is a human protein described in the exemplification of
CC the invention.
XX
SQ Sequence 1681 AA;
Query Match 49.3%; Score 3091.5; DB 23; Length 1681;
Best Local Similarity 52.1%; Pred. No. 1.1e-189;
Matches 656; Conservative 187; Mismatches 274; Indels 143; Gaps 29;
QY 1 MAOFPTPGSLDVAIVVERAKKDOFLSLKPLAGITGDQARNFPFGSLQPOPVLQ 60

```

Db      13 MAQPTAMNGCPNNMATTSEERTKHDRQFNLKPSGGYITGDOARNFELQSLPAPYLAE 72
Qy      61 IVALADNNNDGRMDQVEFFSIAMKLKIKLOGYQLPSTLPVPMKQPV--AISSAPAFGIG 118
Db      73 IVALSDLNKQKMDQVEFFSIAMKLKIKLOGYQLPSTLPVPMKQPV--AISSAPAFGIG 131
Qy      119 GIASMPPLTAAVAPVWG-----SIPVGMSPVLVSVPPAAVPLANGAPV 165
Db      132 ---SMPNLISIPQLPAPAPITLSLSSATSGTNLPMLMPTLVPSPVSTSLP---NGTASL 185
Qy      166 IQPLPAFAHAPATWPKSSSES--RSGGSGQLNTKLOAKS--PDVASAP----- 211
Db      186 IQPLP-IPYSSSTLPKSSSSSLMMGFG--GASIQAKSLIDIGSSSSSTSLSSGNS 241
Qy      212 ---AAEWAVPOSSRLKYLXROLFNSHDKTMSGHLTGPOARTILMQSSLPQAQLASIMNLS 267
Db      242 PKGTSEMAVAPQTRLYKROKFNLTLDKSMGSLSGFARNALQSNLSQQLATIMTLAD 301
Qy      268 IDQDGKLTAEFFILAMHLDVAMSGQPLPVLPEYIPPSFRFRSSGMSVSSSSVDQ 327
Db      302 VDGQGLKAEFFILAMHLDVAMSGQPLPVLPEYIPPSFRFRSSGMSVSSSSVDQ 353
Qy      328 RLPEPSEDEQOEKLPVTFEDKKEENPERGSELEKROALLFOQRKEORLAQLER 387
Db      354 TLPSYQKMG-EFEPQKLPVTFEDKKEENPERGSELEKROALLFOQRKEORLAQLER 412
Qy      388 AEOERKERQOEAKOLELEKOLEKOLEERQEEERKEIERRAPARELEROROLE 447
Db      413 EEMERKOREIOEOMKOLELEKOLEKOLEERQEEERKEIERRAPARELEROROLE 472
Qy      448 WERRROELLNORKEOEGTVLKARRKTLFELEALNDKHOEGLOIRCLATROQ 507
Db      473 WERRROELLNORKEOEGTVLKARRKTLFELEALNDKHOEGLOIRCLATROQ 532
Qy      508 EIESTNKSRELIATHTLOOQOESQOMGLRIPEKOLISDLQKOVO--ONSILHRSLLT 566
Db      533 ELEVLDKQCDLEIMEIKLOQEOYQNKLIYLVPEKOLINERIKNMFQNTSPDSGSL 592
Qy      567 LKRLAKELARQOLREOLDEVERETRSKLOEIVFNNOLEKELHEISKOOLKOROLEA 626
Db      593 HKKSLKEKELC-ORLKEOLDALKEKETASKLEMSFNNOLEKELHEISKOOLKOROLEA 651
Qy      627 ARLKOEKERSLELEK-----EDAOBRVORDKOMLEHVOEOPRPRKHEEDLKR 682
Db      652 KRDKLKEIERKLEIMQKKLEDAARAKAKQKKNENKREKFE-----EKQKRLQE 705
Qy      683 EDSVRKKEAEERAKPEMDKOSRLFHHQEPKALATQAPMSTTEKGLTISAQSVK-- 739
Db      706 EKTQEKIOEERKAEKORDKD-----TLKAEKREKETA 740
Qy      740 --VYVYRALYFERSHDEITIQBGDIYVNDSESQTEGEMLGELKKTGFNPANVAKEXI 797
Db      741 SVLVNRYALYFERNHDEMSFNSGDIIOVEKTVGEEGMVYSGFQNGFPGFNPANVAKEXI 800
Qy      798 P--ENEVETPAKPVTDLTASAPAKLALRETPAPLPVTSSEBSTIPNNWADFFSSTWPSSSN 855
Db      801 PSSENE-----KAVSPKKAL-----LPPTJSLATS-----TSEPELSSN 835
Qy      856 EKPEPTNDWTAAPQSLTVPSAGOLRORSATTPATGSSPYLVGQCEKVEGICQAQALY 915
Db      836 QPASVTDQON--VSFSNLTIVNTSWQ--KKSAPTRTVSPG--SVSPFHGGQVVENIKQAQALC 891
Qy      916 FWRKKNHNLNPNKSDVTVLEQODMMWFGVVOQKGFPSYKLTLSGPRKSTSDTG 975
Db      892 SWTAKGDHNLNPNKSDVTVLEQODMMWFGVVOQKGFPSYKLTLSGPRKSTSDTG 944
Qy      976 PTESBASLKRVAS--PAKPAIPGEEFLAMTYESSSEGGDLTFQOGDVIIVTKKGDWMT 1033
Db      945 -REPEALYAAVNNKPTSAASVGEYIALYPSVSEBGLDTFTEGEBILVTQDGEWMT 1003
Qy      1034 GTVGDKSGVFPNSNYRLDSESGAGTSGSLGKKPEIAQVIATASATATGPQULTLAQOL 1093
Db      1004 GSIGDRSGIFPSNYVKKPKDOESFGSASKSGANNNKPEIAQVITSAVYVSGSEQLSLAQOL 1063

```

```

Qy      1094 IIRKKNPCGMEGBLQARKKROIGWFPANVYKLSPTSKITPTLPKTAOVANVCV 1153
Db      1064 IILKKNPGMWOGELQARKKROIGWFPANVYKLSPTSKITPTLPKTAOVANVCV 1116
Qy      1154 IGMVYTAQNDELAFSGQIIVNLNKEDPDMWKEVSGQVGLFPSSNYVKLTMDPSQ 1213
Db      1117 IAMDYAAANNDELSFSGQIIVNLNKEDPDMWKEVSGQVGLFPSSNYVKLTMDPSQ 1176

RESULT 10
AAV32157
ID AAV32157 standard; Protein: 648 AA.
XX
AC AAV32157;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human SH3D1A protein.
KW SH3D1A; human; Down's syndrome; leukaemia; cancer;
KW megakaryocytic abnormality; myeloproliferative disorder;
KW platelet disorder; neural disorder; thrombocytopenia;
KW haematopoietic disorder; cognitive dysfunction; microcephaly;
KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN WO9953062-A2.
XX
PD 21-OCT-1999.
XX
PR 16-APR-1999; 99WO-US08371.
XX
PR 16-APR-1998; 98US-0082007.
XX
PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX
PI Korenberg JR, Chen X;
XX
DR WPI, 1999-633829/54.
XX
DR N-PSDB; AAZ34573.
XX
PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukaemia.
XX
PS Claim 14; Fig 13; 99p; English.
XX
XS
XX
CC This sequence represents the protein encoded by the human SH3D1A
CC cDNA clone 5 (see AAZ34573). SH3D1A contributes to the development
CC of platelets and the pathogenesis of leukaemias, both in general
CC and in particular those involving the megakaryocytic lineage. The
CC gene maps to the small candidate region for low platelets on
CC chromosome 21. The protein includes SH3 domains and EH domains,
CC both associated with protein-protein interactions and the latter
CC with maintenance of the cytoskeleton. At least 3 isoforms of
CC SH3D1A exist (see AAV32154-58). The invention provides methods for
CC the diagnosis and treatment of megakaryocytic abnormality,
CC myeloproliferative disorder, platelet disorder, acute leukaemia,
CC neural disorders, thrombocytopenia, platelet disorder on
CC chromosome 21, low platelets in deletion for 21, association of
CC gains in chromosome 21 with leukaemias, neural abnormalities,
CC dysfunctions and disorders including brain malformations and
CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
CC and colpocephaly. Methods are also provided for suppressing
CC cells unable to regulate themselves; screening for a somatic
CC alteration in the SH3D1A gene; monitoring the progress and
CC adequacy of a treatment; monitoring tumour risk progress or
CC megakaryocytic abnormality; myeloproliferative disorder,
CC haematopoietic disorder, platelet disorder or leukaemia; and
CC screening of drugs for cancer therapy.
XX

```


SQ Sequence 648 AA:
 Query Match 49.0%; Score 3070.5; DB 20; Length 648;
 Best Local Similarity 94.1%; Pred. No. 6,6e-189;
 Matches 609; Conservative 20; Mismatches 17; Indels 1; Gaps 1;
 QY 1 MAQPTFRFGSLDVAIVTEERAKHDQFLSLKPIAGFITGDDARNFPPQSGIPQVLAQ 60
 DB 1 MAQPTFRFGSLDVAIVTEERAKHDQFLSLKPIAGFITGDDARNFPPQSGIPQVLAQ 60
 QY 61 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGGI 120
 DB 61 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGGI 120
 QY 121 ASMPPLTAAVAVPMGSIPIVVGMSPIVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATVP 180
 DB 121 ASMPPLTAAVAVPMGSIPIVVGMSPIVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATVP 180
 QY 181 KSSSFSGSGSGLNTKLOKAGSFDVASAPPAEMAVPQSSRLKTRQLNNSHDKTMGHL 240
 DB 181 KSSSFSGSGSGLNTKLOKAGSFDVASAPPAEMAVPQSSRLKTRQLNNSHDKTMGHL 240
 QY 241 TGQOARTILMOSSLPOKOLASINWLSIDIDODGKLTAEFTLAMHLIDVAMSGOPLPVP 300
 DB 241 TGQOARTILMOSSLPOKOLASINWLSIDIDODGKLTAEFTLAMHLIDVAMSGOPLPVP 300
 QY 301 PEYIPSPFRFRVSGSGSVISSSVDQRLPEPSSSEDEQP-EKKLPVTFEDKKRENFER 359
 DB 301 PEYIPSPFRFRVSGSGSVISSTSVQRLPEEPVLEDEQOQLEKLPVTFEDKKRENFER 360
 QY 360 GSVELEKRRALLLEQKREDEERLAQLEBAOEKREKREBOEKAKKOLELEKOLEKRELE 419
 DB 361 GNELEKRRALLLEQKREDEERLAQLEBAOEKREKREBOEKAKKOLELEKOLEKRELE 420
 QY 420 ROEEERKKEIEERBAKRELEBOROLEWERNRROELNORNEGSGTVLKARRTLEEF 479
 DB 421 ROEEERKKEIEERBAKRELEBOROLEWERNRROELNORNEGSGTVLKARRTLEEF 480
 QY 480 ELBALNDKKGLEGLQDRIICRLATQROEIESTNKSRELEIAITHLOOOLQESQOMLGR 539
 DB 481 ELBALNDKKGLEGLQDRIICRLATQROEIESTNKSRELEIAITHLOOOLQESQOMLGR 540
 QY 540 LIEKQILSLQKOVQONSILHDSLTLLTKALPAKELARQOLEDEVERETRSKLOEI 599
 DB 541 LIEKQILSLQKOVQONSILHDSLTLLTKALPAKELARQOLEDEVERETRSKLOEI 600
 QY 600 DVANNQKEIREIHSKQOLOKORSLEAARLKOEKOEKRSLEKOE 646
 DB 601 DVANNQKEIREIHSKQOLOKORSLEAARLKOEKOEKRSLEKOE 647
 RESULT 11
 AAM79199
 ID AAM79199 standard; Protein; 1697 AA.
 XX AAM79199;
 AC AAM79199;
 XX 06-NOV-2001 (first entry)
 DT XX
 XX Human protein SEQ ID NO 1861.
 DE XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 OS
 PN MO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001MO-US04098.

XX 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhtman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK52332.
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 4246-4249; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78123-AAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoietic regulatory
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 SQ Sequence 1697 AA:
 Query Match 48.9%; Score 3068; DB 22; Length 1697;
 Best Local Similarity 51.0%; Pred. No. 3,6e-188;
 Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;
 QY 1 MAQPTFRFGSLDVAIVTEERAKHDQFLSLKPIAGFITGDDARNFPPQSGIPQVLAQ 60
 DB 2 MAQPTFRFGSLDVAIVTEERAKHDQFLSLKPIAGFITGDDARNFPPQSGIPQVLAQ 61
 QY 61 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGGI 118
 DB 62 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGGI 120
 QY 119 GIASMPPLTAAVAVPMGSIPIVVGMSPIVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATVP 165
 DB 121 --SWPNTLSIPQPLPAAPPPLTSLSSATSGTLPPLMPTPLVPSVSTSLP---NGTASL 174
 QY 166 IQPLPAFAHPAATVPKSSSFS--RSGPSGLNTKLOKAGSFDVASAP----- 211
 DB 175 IQPLPAFAHPAATVPKSSSFS--RSGPSGLNTKLOKAGSFDVASAP----- 211
 QY 212 ----AAEWAVPQSSRLKTRQLNNSHDKTMGHLGTPARTILMOSSLPOKOLASINWLSID 267
 DB 231 PKTGSSEAVAVQPIRLKXKROKFNLTLSMSGSLGFGARNAHLQSNLSQQLATITWLTAD 290
 QY 268 IDQDGLTAEFTLAMHLIDVAMSGOPLPVPPEYIPSPFRFRVSGSGSVISSSVDQ 327
 DB 291 VDQDGLTAEFTLAMHLIDVAMSGOPLPVPPEYIPSPFRFRVSGSGSVISSSVDQ 342
 QY 328 RLPEPSSSEDEQOPEKKLPVTFEDKKRENFERGVELEKRRALLLEQKREDEERLAQLE 387
 DB 343 TLPSYQKMG-EEEPQKLPVTFEDKKRENFERGVELEKRRALLLEQKREDEERLAQLE 401

```

QY 388 AEOERKEREPOEAKOLELEKOLEKORELEFORBERKERKEIEREDAKRELEFOROLE 447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 402 EEWERKORELEOEBWKOLELEKOLEKORELEKOREBEREKDEREBAKOLEOROLE 461
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 448 WEENROELNONKKEGTVLKARKKTELEFELEALNDKHOLEGGLOIRCLLATORO 507
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 462 WERROELNONKKEGTVLKARKKTELEFELEALNDKHOLEGGLOIRCLLATORO 521
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 508 EISTNSKRELEIAETITLLOOLESOOMLGRLLPEKOLISDOLKOVQ-ONSILHSDSLT 566
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 522 ELEVLDRKOCOLEIMEIKOLEOLEYONKLLIYVPEKOLINERIKMNOFNTDSDGVSLL 581
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 567 LKRALEKELARQOLROLDREVERETSKEIETDVFNNO----- 605
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 582 HKKSLEKEELC-ORLKEOLDLEKETASKLESDMSFNNOCKGNDSDVLCILLISLCL 640
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 606 -----LKEIREISHKOOLOKORSLAARLKOKEOERKSLELEKOK-----EDAORVOER 655
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 641 NNLFLLKELERETYNTOQLALEOLYKIKRDKLEIERKRELMOKKLEDEAKKAKKOGK 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 656 DKOWLEHVQOEEOPRPRKPHEDRLKREDSVRKKEAEERAKPEMOKOSRLEFPHOEPK 715
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 701 ENLWKENLRKEE-----EKOKRLOEKTOKELQEBERKAEERKOKD----- 744
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 716 LATQAPWSTTEKGPLTISAQESVK-----VYVYRALYPPFSRSHDETTIQPDIVWDES 770
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 745 -----TLKAEKKREKTASVLNVRALYPERARNDEWSNDIIOVDEK 789
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 771 QTEGPGWLGELKGTGKTFPANYAEKIP--ENEVYTPAKPVTDLTSAAPKALRETAP 828
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 790 TVSGPGWLYSGFQGNFGMPCNVYKMPSSENE-----KAVSPKAL----- 831
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 829 LPVTSSBPTPLNNWADPSSSTWPSSENEKEPTDNWMAOPSLTVSAQOLKORSAFTP 888
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 832 LPPTVLSLATS-----TSSEPLSSNQPPASVTDYQV-VFSNLTNTVTSWO--KKSAPTR 881
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 889 ATATGSSPSPLVAGGEGEVLQALYPMWAKKDNHLPNKSDVITVLEQODMMWEGEVQ 948
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 882 TVSGG-SVSPFHGGGOVENKALCSWTRAKDNHLPNFKHDITVLEQODMMWEGEVH 940
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 949 GQKGMFPKSYVKLISGVRKSTSIDTPTSPASLKRVAS--PAKPAITGEEFIAMTYT 1006
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 941 GGRGMFPKSYVKIIPGSEVK-----REPEPALYAAVKKPPTSAVSGEETIALY 992
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1007 ESSGQGLTPOGGDVIVYTKKDGDMWTGVDKSGVPSNVVRLKDESGGTGKTGSLG 1066
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 993 SSVPGDLTFEGEELVTQKDGEMWTGSDGRGIFPSNVVFKDDESGSASKSASN 1052
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1067 KKEPIAOVIASVATGPEQLTLAPGOLILIRKKNPGMWEGELQARGKROIGMFPANY 1126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1053 KKEPIAOVTASVVASGSEQLSLAPGOLILIKKNTSGMWGELQARGKROKGMFPASHV 1112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1127 KLSFGTSKITPTLEPTAVOAVCYIGMDYTAONDDELAFSGOITIVLAKEDPDW 1186
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1113 KLDPSSERRATPAHP-----VCQIAYMYDANNEDLSFSKQILINVMKDDPDW 1165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1187 KGEVSGGVGLPSPNVYKLTITMDPSQO 1213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1166 QGEINGVTGLPSPNVYKLTITMDPSQO 1192
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KM antiproliferative; antiviral.
XX Mus sp.
XX MO9955728-A2.
XX 04-NOV-1999.
XX 27-APR-1999; 99MO-CA00375.
XX 27-APR-1999; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX (HSCR-) HSC RES & DEV LP.
XX Egan SE, Mang W, Sengar A;
XX WPI; 2000-052802/04.
XX N-PSDB; AAZ39010, AAZ39011.
XX New nucleic acid encoding Ees2 and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection.
XX Claim 33; Page 48; 99pp; English.
XX The present sequence represents mouse Ees2. The present invention
XX specifically describes mammalian Ees1 and 2 proteins (I) and their splice
XX variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
XX are involved in regulation of clathrin-mediated endocytosis (as a complex
XX with Esg1 protein), vesicular trafficking and actin cytoskeleton.
XX Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
XX mutants), (I)-specific antibodies (Ab); sequences antisense to the (I)
XX polynucleotide; agents that downregulate expression of Ese genes or
XX antagonists of an Ese binding partner are used to treat diseases
XX associated with undesirable endocytosis and resulting changes in cellular
XX function. Particularly overexpression of Ees1 is used to block
XX clathrin-mediated endocytosis in vivo or in cell cultures, while
XX administration of (I) is used to promote endocytosis of selected cells.
XX (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
XX cells that can be stimulated to proliferate by a growth factor receptor;
XX and similar compounds (also inactive Ese mutants) can be used to prevent
XX viral infection. Endocytosis may also be regulated, in vivo or in cell
XX cultures, by forming an Ese-Esg1 complex, then binding dynamn to the
XX complex. Generally conditions that can be treated include cancer;
XX abnormal cell division or migration; viral infection; or abnormal
XX receptor signalling, tissue development or synaptic transmission.
XX
XX Sequence 1197 AA;
SQ
Query Match 48.6%; Score 3049.5; DB 21; Length 1197;
Best Local Similarity 52.4%; Pred. No. 3.5e-187;
Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
QY 1 MAQPTTPEGSLDVAITVEERAKHQDQLSLKPIAGFITDQARNFPFGSLPQVLAQ 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MAQPTTANMGNMMAITSEERTKHQDFNLKPSGVIITDQARTFLOGLPAPVLA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 IVALDMMNDGMDQVRSIMKLIKXKLGQVLPSTLPVYMKQOPV--ATSSAPARGIG 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IVALSDLNKMDQDQERSIAMKLIKXKLGQVLPSTLPVYMKQOPV--ATSSAPARGIG 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 119 GIASMP-----PLTAVAPV--PMG-----SIPVGMSPPLVSVPPAAVPLANGAP 164
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 ---SMPNLSIHQPLPVPVPIATPLSSATSGTSLPMLMMPALVSVSTSLP--NGTAS 173
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 165 VIQPLPAPAPAPATWPKSSSS--RSGGSQLNTKLQQAQ--FDVYASAP----- 211
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 174 LIQPL-SIPYSSSTLPHASYSIMMGFG---GASIQRAQSLIDLGSSTSTSTASLSGN 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 212 -----AAEWAVQSSRLVRYROLFNSHDKTWSGHLTGPAFTILMOSSLPQQLASIMNLS 266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 12
 AAY57445
 ID AAY57445 standard; Protein: 1197 AA.
 AC AAY57445;
 XX 28-FEB-2000 (first entry)
 DT Mouse Ees2 protein sequence.
 DE Mouse Ees2 protein sequence.
 XX Mouse; murine; Ees1; Ees2; endocytosis; vesicular trafficking;
 KM regulation; actin cytoskeleton; detection; cancer; infection;

```

Db 230 SPPTGSENAVPOPSRLKTRQKFNSLDKMGVLSGFGQARNALQSLNLSOTLATITWLA 289
Qy 267 DIDODGLTAEEFIILAMHLDVAMSGQPLPVLPPEYIPPSFRVSSGMSVSSSYD 326
Db 290 DIDGDLQALAEFIILAMHLDMAKAGQPLPLTIPPELVPSFR-----GGKV---DSYN 341
Qy 327 QRLPESSSEDEQPEKPLPVTFEDKKRENFERSVELEKRRQALIEQQRKEQERLAQLE 386
Db 342 GTLPYQKTO-EEEPQKLPVTFEDKKRKNYERGMNELKRRQVLMEEQQQRAERKAQKE 400
Qy 387 RAEQEKREERQEAQKQLELEKQELERQREERREKEIEREAKELERQRL 446
Db 401 KESEKREKQRELEQEWKQLEKLEKQELERQREERREKEIEREAKQLEERQRL 460
Qy 447 EMERNRQELLNQRNKEQGTVLKARKTLFEFLALNDKXHQLEGGKQDRLCRILATOR 506
Db 461 EWERLARQELLQKTEQEDIVRLSRKSLHLELEAVNGKHQISGRQDVQIRKQKOT 520
Qy 507 QEIESTNKRRELIATITHLQOQLOESQOMLGRLEPKQILSDQLKQVQNSLHRDSLIT 566
Db 521 TELEVLDKQDLEIMEIKOLOQELKEYQNKLIYVPEKQOLLNERIKNMQLSTNTPSGISL 580
Qy 567 LKRALEAKELARQALEQLEDEVRETRSKQLEIDVNNQLEKEIRHSQOQOKQRSLEA 626
Db 581 LHKSSSEKELCQRLKEQDLEKETAASKLSEMDSPNNQLEKELRESYNTQDLALQELHRI 640
Qy 627 ARLKQKEQERKSL-LEKQK---EDAQRAVERQDKQMLHVOOEOPRPRKHEEDRLK 682
Db 641 KRDQKLEIRKRLQJOYKQKLEDEAKAKQKQENMRSEIRKEE-----EKQRLQOE 694
Qy 683 EDSVRKKEAEERAKPEMOKQSLFHHQEPAPKATQAPWSTTEKGPLTISAQESVAVVY 742
Db 695 EKSQDTQEEER---KAEAKQSE-----TASA-----LVN 721
Qy 743 YRLAYFEERSHDEITIQGDIYMWDESOTGERGWLIGELKQKTFMFPANVAKETPENV 802
Db 722 YRLAYFEERNHDEMSFGSDIIOVEKTVGEPGWLVSFGQKFGFPCNVYEVKVLSE- 780
Qy 803 PTPAKPVTLTSPAPKALARETPAPLPVTSSPPTPNMADFSTWSSSEKRETON 862
Db 781 -----KALSPKALLPPTVLSLSTSS-SQPPASVTDYAHV--SFSNLVTNT-- 824
Qy 863 WDTWAAQSLTVPSAQQLRQSAFTPATATGSSPSPLVQGEKEVGLQAQALYPMRAKQD 922
Db 825 --TW-----QOKSAFTRTVSPG-SVSPIHQGGQAVENLKAQALCSMTAKKE 867
Qy 923 NHIENPKSDVITVLEQDDMMWFEVVGQKGMFPKSYVXLISC-PVAKS--TSIDGPTES 979
Db 868 NHIENFKHDVITVLEQDDMMWFEVVGQKGMFPKSYVXLIPGNEVQRGEPALYAAVTCK 927
Qy 980 PALKRVAAPAKPAIP-GEEFIAMTYTESSEGGDLTPQGGDVIVTKKDGGMWGTGVD 1038
Db 928 PIS---TAIPVISTATVPEGEDIALYSYSVEFGDLTFEGEILVLTQDQGEWMTGSIE 984
Qy 1039 KSGVPSNVYRLKDSBEGTAGTGLGKKPELAIQVIAQVIAATGPBQULTLABQULIRK 1098
Db 985 RTGIFPSNVYRPRDQGNFNGASKSGASNNKPELAIQVIAASGTRQLSLAQULILIK 1044
Qy 1099 KNPGMWEGELQARKKQKRGQIGWFPANYVVLKLSFGTSKITPTLEPKTAVNPANQVIGMD 1158
Db 1045 KNTSGMWQELQARKKRGKRGWFPASHVXLKLGSSSERTPT-----PHAQCOVIAMD 1097
Qy 1159 YTAQNDDELAFSGQIITVNLNKEDPPMWKQEVGQGVGLFPPSNVYVLTMTDPSQO 1213
Db 1098 YMANNEDELNFSQGLINWKNKDDPDMWQGTNGLTGLFPPSNVYVMTTSDPSQO 1152

```

```

RESULT 13
AA57450
ID AA57450 standard; protein; 1658 AA.
AC AA57450;
XX

```

```

DT 28-FEB-2000 (first entry)
XX
Db Mouse Ese2L protein sequence.
XX
KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
OS Mus sp.
XX
XX M09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WC-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
XX
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Egan SE, Wang W, Sengar A;
XX
DR WPI; 2000-052802/04.
XX
DR N-PSDB; AA239026, AA239027.
XX
PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection.
XX
PS Claim 33; Page 69-70; 99pp; English.
XX
CC The present invention specifically describes mammalian Ese1 and 2
CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-
CC mediated endocytosis (as a complex with Eps15 protein), vesicular
CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC sequences antisense to the (I) polynucleotide; agents that downregulate
CC expression of Ese genes or antagonists of an Ese binding partner are
CC used to treat diseases associated with undesirable endocytosis and
CC resulting changes in cellular function. Particularly overexpression of
CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
CC cultures, while administration of (I) is used to promote endocytosis of
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC proliferation of cells that can be stimulated to proliferate by a growth
CC factor receptor; and similar compounds (also inactive Ese mutants) can be
CC used to prevent viral infection. Endocytosis may also be regulated, in
CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
CC dynamin to the complex. Generally conditions that can be treated include
CC cancer; abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission. The
CC present sequence represents mouse Ese2L protein sequence.
XX
SQ Sequence 1658 AA;

```

```

Query Match 48.6%; Score 3049.5; DB 21; Length 1658;
Best Local Similarity 52.4%; Pred. No. 5; se-187;
Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;

```

```

Qy 1 MAOPFPFGGSLDVAVTVEERAKHQDFLSKPIAGFITGDDARNFPPQSGHPQPLAQ 60
Db 1 MAOPFTMANGSPNNMALTSEERTKHDQFNLKRSQGYITIGDQARTFLOSGLPAPVLA 60
Qy 61 IVALADNNDRMDQVEFSEIAMKLIKUKLQGYQPLSTPLPVMKQQPV--AISSAPAFGIG 118
Db 61 IVALSDLNKQKMDQGFSEIAMKLIKUKLQGOQLPVVLPIMKQPFWSPLISA-RFGMG 119
Qy 119 GIASMP-----PLTAVAPV--PWG-----SIPVGMSPVLSSVPPAPVPLANGCAP 164
Db 120 ---SMPNLSTHQPLPVAPVPIATPLSSATSGTSLPPLMMPAPLVPSVSTSLP--NGTAS 173

```

```

QY 165 VIOLPAPAPAPATWPKSSSES--RSGPSQNTKLQKAS--FDVASAP----- 211
DB 174 LIQPL--SIPISSSTLPIASSYSLMGFG--CASIQKASOLIDLSSSTSTASLSCN 229
QY 212 -----AAEMAVPOSSRLKYROLFNSHDKTMSGLTGPQARTILMOSSLPQAOIASIWNLS 266
DB 230 SPKTGSEMAVPOPSRLKYRQKFNSLDKMGSGYLSGFOANALLQSNLSQTOIATITWLA 289
QY 267 DIQDGLTLEEFILMHLIDVAMSGOPLPVLPPETIPPSFPRVSSGSMYSISSSD 326
DB 290 DIGDGLKKEEFLMHLIDVAMKAGQPLPLPPELVPSF-----GKQV--DSVN 341
QY 327 QRLPEEBSSEDOPEKPLPTFEDKKRENFERSGVELEKROALLQORKEQERLAQLE 386
DB 342 GLPVSQKTL--EERPKKLPVTFEDKKKAYERGNMELERKQVLMWEOQREARAKQKE 400
QY 387 RAEQERREERQOEAKROULEKOLEKOREROREEERKEITEREAKRLEEROROL 446
DB 401 KEEMERKORLEOQEMWKOLELEKRELEROREEERKEITEREAKOLEEROROL 460
QY 447 EEMERNROELNORKEOEGTVLKARKTLFELEALNKKKOLBGKLDIDICRLATOR 506
DB 461 EEMERLROELLSQKTRQEDIVLSSRKSLHLEAVNKHQOIGRLQDVQIRKQTK 520
QY 507 QEISTNKSELRILAEITHLQOOLQESQOMLGRILPEKQILSDQKOVQONSILHRSLLT 566
DB 521 TELEVLDRKODLEIMEIKOLEKEKONKLIYLVPEKQILNERIKNMQSLNTPDSGISL 580
QY 567 LKALAEKELAROOERLEVERETRSKLOEIDVFNNOLKEIREHSKQOLQORSLLEA 626
DB 581 LHKSSSEKEICORLQKQDLEKETASKLSEMDSPNQKELRESINTQOLLEQHLKI 640
QY 627 ARUKOEKERSLE--LEKOK---EDAQRVOERDKOMLEHVQOEOPRPRKHEEDRLKR 682
DB 641 KRDKLKEIERLEJOIKKLEDEAKKAKQKQENLWRESIRKEE-----EKORLOE 694
QY 683 EDSVRKKEAEERAKPEMODQOSRLFHPHOPAKLATAQAPSTTEKGLTISAGESVKVVY 742
DB 695 EKSQDKTOEER--KAERAKQSE-----TASA-----LVN 721
QY 743 YRALPEESRSHDEITTOPGDIWVWDESQTEPGMLGELKQKTMFPANVAKRIPENEV 802
DB 722 YRALYPEANRHDMSFSSGDIIOVDEKTVGEPGMLVSGQKFGMPCNVYEVKVLSE-- 780
QY 803 PTPAKPVTLTSAAPKALARETAPLPVTSSEPTPTNNMADPSSSTWSSSEKRETDN 862
DB 781 -----KALSPKALLPPTVLSATSTST--SOPPAVTDVHNV--SFSNLTVNT-- 824
QY 863 WDTWAAQPSLTVPSAGOLRORSATTPATATGSSSPVLAGQEKVEGLOAOLYPMRAKQD 922
DB 825 --TW-----QKSAFTRIVSPG--SVSPHGGQAVENLKAQALCSMTAKKE 867
QY 923 NMLNFKSDVITVLEQODMMWFGVQOKGMPKSYKYLISG--PYRKS--TSIDTGTES 979
DB 868 NMLNFKSDVITVLEQODMMWFGVHGRGMPKSPKSYKYLIPGNEVORGEDEALYAATTK 927
QY 980 PASLKRVAAPAKAIP--GEFIMATYTESSEOGDLTFQOQDVIVVTKXGDMWTGVGD 1038
DB 928 PTS---THPYPTSTAYPVGEDYIALYSSVPEPDLFTTEGEBILIVQKQGMWTSIGE 964
QY 1039 KSGVFPSSNYVRLKDSGSGTAGTKSLGKKPEIAQVIASVYATGPEQLTAPQOLIRK 1098
DB 985 RTGIFPSNYVAPKQOENGNASKSGASKKPEIAQVITSAVAGSTEQLSIAPQOLILIK 1044
QY 1099 KNPBGWMEGELQARQKQIQGMFPANVYKLLSPGSKITPPELPTKAVOPAVOCVGMVD 1158
DB 1045 KNTSGMWQEGELQARQKQKQKMPFASHVKLLGPPSSERTMP-----FHAVOCVIAKMD 1097
QY 1159 YTAQNDDELAFASQOILVNLKEDPDMWKGEVSGQVGLFPSNYVYKLLTTDMPSQO 1213
DB 1098 YMANNEDELINFSKQOILVNMKDDPDMWQGETNGLTGLFPSNYVYKMTTDSQPSQO 1152

```

```

RESULT 14
AA71160
ID AA71160 standard; Protein; 1683 AA.
XX
AC AA71160;
XX
DT 08-SEP-2000 (first entry)
XX
DE Rat phosphodiesterase interacting protein, M14.
XX
KW Rat; phosphodiesterase interacting protein; M14; PDE; CAMP-PDE;
KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
KW antiinflammatory; antispasmodic; dermatological; antibacterial; shock;
KW analgesic; immunosuppressive; antitumor; vasodilator; antiarthritic;
KW antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;
KW antineumatic; treatment; inflammatory disease; psoriasis; arthritis;
KW atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
KW eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
KW reperfusion injury; atopic dermatitis; diabetes insipidus;
KW conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
KW arterial restenosis; ankylosing spondylitis; transplant rejection;
KW graft versus host disease.
XX
OS Rattus sp.
XX
PN WO200027861-A1.
XX
PD 18-MAY-2000.
XX
PE 12-NOV-1999; 99MO-US26860.
XX
PR 12-NOV-1998; 98US-0108255.
XX
(STRD ) UNIV LEIAND STANFORD JUNIOR.
XX
PI Conti M, Pahlke G;
XX
DR WPI, 2000-376479/32.
XX
PT Polynucleotide encoding a phosphodiesterase (PDE) interacting
PT polypeptide, useful for diagnosis and treatment of asthma, cystic
PT fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
PS Disclosure; Fig 6; 77pp; English.
XX
CC The present sequence is a phosphodiesterase (PDE) interacting protein,
CC M14 from rat. The protein modulates the functions and properties of PDEs,
CC specifically CAMP-PDEs, and also targets them to specific subcellular
CC compartments. The present sequence
CC can be used in the diagnosis and treatment of disease conditions
CC associated with PDE activity. The diseases include asthma, cystic
CC fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
CC granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
CC septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
CC inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
CC distress syndrome, diabetes insipidus, allergic rhinitis, allergic
CC conjunctivitis, vernal conjunctivitis, arterial restenosis,
CC atherosclerosis, inflammatory diseases associated with irritation and
CC path, rheumatoid arthritis, ankylosing spondylitis, transplant
CC rejection and graft versus host disease, disease conditions associated
CC with hypersecretion of gastric acid, and disease conditions in which
CC cytokines are mediators.
XX
SQ Sequence 1683 AA;
XX
Query Match 48.6%; Score 3046; DB 21; Length 1683;
Best Local Similarity 50.8%; Pred. No. 9,4e-187;
Matches 652; Conservative 187; Mismatches 267; Indels 178; Gaps 33;
QY 1 MAQPTPFGSGLDVVAITVEERAKHQDQFLSLKPIAGFITGDDQANFPFGGLPQVLAQ 60
DB 2 MAQPTANMGGNMMAITSEERTKHQDFNLKPSGQYITGDOARTFLOGLPAPVLA 61

```

```

QY 61 IVALADNMNDGRMDQVEFSIAMKLILKLOGYOLPSTLPPVMQOV--AISSAPFGIG 118
DB 62 IVALADNMNDGRMDQVEFSIAMKLILKLOGYOLPSTLPPVMQOV--AISSAPFGIG 120
QY 119 GIASMP-----PUTAAVAV--PMG-----SIPVGMSPPLVSPVPAVPLANGAP 164
DB 121 ---SMRPLSHOPLPPVAPITAPLSSATSGTSLPPLMPAPLVPSVSTSSLP---NGTAS 174
QY 165 VIQPLPAFAHAPATWPKSSFS--RSGPGSOLNTKLQAKOS--FDVASAPP----- 211
DB 175 LIQPL-SIPYSSSTLPHASSYSIMMGFG--GASIQAKOSLIDLSSSTSTASTASG 230
QY 212 -----AAEMAVPOSSRLKRLQULFNSHDKTMSGHLTGPQARTIMOSLSLQOALASIMNS 266
DB 231 SPKGTSEMAVVPQPSRLKRLQULFNSHDKTMSGHLTGPQARTIMOSLSLQOALASIMNS 290
QY 267 DIQDQKLTKEEFLIAMLHLDIVAMSGOPLPVLPEXIPSPFRVAVSGSMVSISSSD 326
DB 291 DIQDQKLTKEEFLIAMLHLDIVAMSGOPLPVLPEXIPSPFRVAVSGSMVSISSSD 342
QY 327 QRLPEEPPSEDEOQPEKULPVTFEDKKRENFERGVSYLEKRRQALLQOQKEOERLAQLE 386
DB 343 GLTLPYQKQKQ-EEEPQKQPVTFEDKKRENFERGVSYLEKRRQALLQOQKEOERLAQLE 401
QY 387 RAOEKERERROQEAQKOLEKOLEKORELROREBERRKEIFEREAQKLEERQOL 446
DB 402 KEEMERKORLEOQEMKOLEKOLEKORELROREBERRKEIFEREAQKLEERQOL 461
QY 447 EMERNROELNORNKOEGBTVLKAARKTLEFELALNDKHKQLEGLKODIRCLATQR 506
DB 462 EMERIRROELNORNKOEGBTVLKAARKTLEFELALNDKHKQLEGLKODIRCLATQR 521
QY 507 OEIESTNKSRELIAETILHLOOQLOESQOMLGRILPEKOLISDOLKOVONSLHRDLSLT 566
DB 522 TELEVLADKQCDLIMEIKOQLOQLOEYONKLIYLPBEKOLNRIKQMLNSTPDSGIBL 581
QY 567 LKALAEKELARQOLBOLDEVERETRSKLOETIDVFNNO----- 605
DB 582 LHKSESEKELCORLEKQDLEKELTASKEEMDSFNQOLKQNMDSVLOCLISLSLCL 641
QY 606 -----LKELEHSHKQOOLOKORSLKLEARKLQKOEKRSLE-LEKOK---EADQRRVQR 655
DB 642 NMLFLILKEIRBESYNTQOALBOLHKKIKRDKLELKRLEJOQKLEDEARKAKQKQK 701
QY 656 DKQWLEHVQOEOPRPRKPHEDRLKREDSVRRKEAEERAKPEMOKOSRLFFHQEPAPK 715
DB 702 ENLMKRSIRKEE-----EKQKRLQEKSOQDRTQEBR----- 734
QY 716 LATQAPMSTTEKCPITISAOESVAVVYRALYPPESRSHDEITIQODIWMVDESQTEP 775
DB 735 -KTEAKQSEETARA-----LVNYRALYPPFARNHDEMFSNGDIIQVDEKTVGEP 782
QY 776 GWLGGELKGTGWFANVYAEK-IPENEVPTPAK---PVTDLISAPKALARETAPILP 830
DB 783 GWLYGSPQGFQFPCNVEKMLSDKTBSPKALLPRAVLSATSA-----APQP 833
QY 831 VTSSBSTTPNNMADFSSTWSSSNEKPEITDMDTAAQPSLTVPSAGOLRQSAFTPAT 890
DB 834 LGSNQAAPV---TDYQNV---SFSNLAVNT---TW-----QOKSATRTIV 869
QY 891 ATGSSSPSVLQGEKAVGLQOALYPMRAKKDNLNPNKSDVITVLEQDDMMWFGVQOQ 950
DB 870 SPG-SVSPHIGQOQAVENLKAQALCSWTAKKENHNFPSKHIDVITVLEQOENWVFGVHG 928
QY 951 KGMFPKSVYVLIGSPVARKSISIDTGPES-PASLKRVASAPALPGEFTIAMTYEES 1009
DB 929 KGMFPKSVYVLIG-----SEVRKGEPEALYAAVANKPSTAPV--GEVYIALYSYSSV 981
QY 1010 EOODLTFQOQDVLIVTKDDMTGTGVDKSGVFPANVYALKDSSESGTAGTKSGKRP 1069
DB 982 EPDPLTFTEBEELLVYQKXOEWMTSGISGERTGIFPSNYVRPKQDEVANASGASANKRP 1041
QY 1070 EIAQVIASTVATGPBOLTAPGOLILIRKKNPGGWEGELQARGKKRQIGWFPANYVLL 1129

```

```

DB 1042 EIAQVIASTVATGPBOLTAPGOLILIRKKNPGGWEGELQARGKKRQIGWFPASHVLL 1101
QY 1130 SPGSKITPTELPATAVPAVCAQVIGWYDYAONDDLEAFSGQIINVLNKEDPDMMKGE 1189
DB 1102 GPSARTTP-----AFHACVQIAMTYDYIANNEDELNFSKGLINVANKDPPDMQGE 1154
QY 1190 VSGQVGLFPPSNVYKLTDTMDPSQ 1213
DB 1155 INGVGLFPPSNVYKLTDTMDPSQ 1178

RESULT 15
AAM93229
ID AAM93229 standard; Protein; 642 AA.
AC AAM93229;
XX
XX
DT 06-NOV-2001 (first entry)
DE Human polypeptide, SEQ ID NO: 2647.
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
XX Homo sapiens.
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H,
DR WPI: 2001-524255/58.
DR N-PSDB; AAK94139.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2647; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 642 AA;

Query Match 48.4%; Score 3034.5; DB 22; Length 642;
Best Local Similarity 93.9%; Pred. No. 1,3e-186;
Matches 603; Conservative 18; Mismatches 20; Indels 1; Gaps 1;

QY 1 MAQFPTPGGSLDWAITVERAKGDOQFLSLKPIAGFITGDOARNFPOSGLPQVLAQ 60
DB 1 MAQFPTPGGSLDWAITVERAKGDOQFLSLKPIAGFITGDOARNFPOSGLPQVLAQ 60
QY 61 IVALADNMNDGRMDQVEFSIAMKLILKLOGYOLPSTLPPVMQOVPAISSAPFGIGI 120

```

```

Db 61 IWLADANNNOQRMQVFEFIAIMKLIKLIKQYQLPSALPVMKQOPFAISSAPAFGAGI 120
QY 121 ASMPPLTAVAPVPMGSTIPVGMSPPLVSVPPAVPPLANGAPPVIOPLPAFAHPAATWP 180
Db 121 ASMPPLTAVAPVPMGSTIPVGMSPPLVSVPPAVPPLANGAPPVIOPLPAFAHPAATWP 180
QY 181 KSSSFSSRSGSQSLNTKLOKQAFDVAASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHL 240
Db 181 KSSSFSSRSGSQSLNTKLOKQAFDVAASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHL 240
QY 241 TGPOARTILMOSSLPOQOLASIMWLSDIDOGKLTAEFFILAMHLIDVAMSGQPLPVLP 300
Db 241 TGPOARTILMOSSLPOQOLASIMWLSDIDOGKLTAEFFILAMHLIDVAMSGQPLPVLP 300
QY 301 PEYIPPSFRVRSGSGSVISSSVSDORLPEPESSEDEOOP-EKULPVTEDKKRENFER 359
Db 301 PEYIPPSFRVRSGSGSVISSSVSDORLPEPESSEDEOOP-EKULPVTEDKKRENFER 359
QY 360 GSVELKKRQALBEOQRKEQERLAQLERABOERKEREROQAKROLELEKORELE 419
Db 361 GNLLEKKRQALBEOQRKEQERLAQLERABOERKEREROQAKROLELEKORELE 420
QY 420 ROEEERKEIERREAKRELEROROLEMERNRROELLNOANKOEGTVVLKARKTLEF 479
Db 421 ROEEERKEIERREAKRELEROROLEMERNRROELLNOANKOEGTVVLKARKTLEF 480
QY 480 ELEALNDKKHOLEGKLODIRCRLATOROLESTNKSRELRIAEITHLQOOLQESQOMLGR 539
Db 481 ELEALNDKKHOLEGKLODIRCRLATOROLESTNKSRELRIAEITHLQOOLQESQOMLGR 540
QY 540 LIPEKOIISDOLKOVOONSJHRDSJLTIKRALAKELARQOLREQLDVEVERETRSKLOEI 599
Db 541 LIPEKOIISDOLKOVOONSJHRDSJLTIKRALAKELARQOLREQLDVEVERETRSKLOEI 600
QY 600 DVFNNOJKELEIRHSKOOLOKORSLEAARLKOEKERSLEL 641
Db 601 DVFNNOJKELEIRHSKOOLOKORSLEAARLKOEKERSLEL 642

```

Search completed: March 14, 2003, 11:37:45
 Job time : 79 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:37:26 ; Search time 476 Seconds
(without alignments)
117.457 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6263
Sequence: 1 MAQFPTPGGSDVMAITVE.....VGLFSPSNVYKLTMDPSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058.5	32.8	462	10 US-09-879-957-38	Sequence 38, Appl
2	1388.5	22.1	509	10 US-09-879-957-194	Sequence 194, App
3	1330.5	21.2	530	9 US-09-764-868-738	Sequence 738, App
4	816.5	13.0	248	10 US-09-879-957-40	Sequence 40, Appl
5	527	8.4	813	9 US-09-864-899-25	Sequence 25, Appl
6	429	6.8	90	10 US-09-864-761-33620	Sequence 33620, A
7	429	6.8	90	10 US-09-864-761-43004	Sequence 43004, A
8	426	6.7	99	10 US-09-864-761-31635	Sequence 31635, A
9	422	6.7	82	10 US-09-864-761-46565	Sequence 46565, A
10	413	6.6	87	10 US-09-864-761-34169	Sequence 34169, A
11	325	5.2	659	9 US-10-144-621-2	Sequence 2, Appl
12	322	5.1	59	10 US-09-879-957-133	Sequence 133, App
13	316.5	5.0	287	9 US-09-764-868-744	Sequence 744, App
14	314	5.0	62	10 US-09-879-957-135	Sequence 135, App
15	309	4.9	56	10 US-09-879-957-114	Sequence 114, App
16	303.5	4.8	611	10 US-09-216-393-81	Sequence 81, Appl
17	300.5	4.8	1175	10 US-09-771-161A-224	Sequence 224, App
18	300.5	4.8	1175	10 US-09-771-161A-225	Sequence 225, App
19	300.5	4.8	1175	10 US-09-771-161A-226	Sequence 226, App

20	296	4.7	57	10 US-09-879-957-136	Sequence 136, App
21	293	4.7	888	9 US-09-893-519A-73	Sequence 73, Appl
22	288	4.6	59	10 US-09-864-761-33618	Sequence 33618, A
23	287.5	4.6	76	10 US-09-864-761-34168	Sequence 34168, A
24	287.5	4.6	82	10 US-09-864-761-46614	Sequence 46614, A
25	287.5	4.6	350	9 US-10-101-487-58	Sequence 58, Appl
26	285.5	4.6	830	9 US-10-033-245-7	Sequence 7, Appl
27	285.5	4.6	830	9 US-10-033-223-7	Sequence 7, Appl
28	285.5	4.6	830	9 US-10-033-167-7	Sequence 7, Appl
29	285.5	4.6	830	9 US-10-033-244-7	Sequence 7, Appl
30	285.5	4.6	830	9 US-10-033-435-7	Sequence 7, Appl
31	285.5	4.6	830	9 US-10-032-990-7	Sequence 7, Appl
32	285.5	4.6	830	12 US-10-033-246-7	Sequence 7, Appl
33	285.5	4.6	830	12 US-10-033-301-7	Sequence 7, Appl
34	285.5	4.6	830	12 US-10-033-326-7	Sequence 7, Appl
35	281	4.5	788	10 US-09-879-957-30	Sequence 30, Appl
36	277	4.4	97	10 US-09-864-761-36740	Sequence 36740, A
37	277	4.4	101	10 US-09-864-761-37838	Sequence 37838, A
38	275	4.4	103	10 US-09-864-761-36224	Sequence 36224, A
39	275	4.4	1360	10 US-09-871-916-2	Sequence 2, Appl
40	273.5	4.4	1005	10 US-09-925-301-1335	Sequence 1335, Ap
41	271	4.3	88	10 US-09-864-761-36232	Sequence 36232, A
42	270.5	4.3	2568	10 US-09-866-108-3	Sequence 3, Appl
43	264	4.2	1805	9 US-09-820-843A-73	Sequence 73, Appl
44	264	4.2	1945	9 US-09-927-597-2	Sequence 2, Appl
45	264	4.2	1979	9 US-09-927-597-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-879-957-38
Sequence 38, Application US/09879957

Patent No. US20020034755A1

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

HOFFMAN, No. US20020034755A1h

KAY, Brian K.

FOWLER, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879, 957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630, 915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistic, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 38:

; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: 60/236,893
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/298,309
 ; PRIOR FILING DATE: 2001-06-14
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 25
 ; LENGTH: 813
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-664-899-25

Query Match 8.4%; Score 527; DB 9; Length 813;
 Best Local Similarity 21.9%; Pred. No. 3e-18;
 Matches 197; Conservative 138; Mismatches 283; Indels 280; Gaps 26;

QY 20 EERAKHQOQFLSTKPIAGFITGDOARNFPOGSLPOPVLQOIMALAMNMDGRDQVEFS 79
 DB 105 EDKAKYDAIFDSLSPVNGFSGDKVPVLSKLPLVDILGRWELSDIDHGMIDRDEFA 164
 QY 80 IAKMLIKLQGYOLPSTLPVPMKQOYVAISSAPAFGIGIASPPLTAAVAPVPMGSI PV 139
 DB 165 VAMFLVCALEKEBPVMSLPPAL---VPSEKRTVSGISGVRLIPSSAKESYHSLPS 220
 QY 140 VGMSPPLVSVPPAAVPLANGAPVVIQPLPAFAHPAATPMKSSFSRSGSQNLTKQ 139
 DB 221 VGILP----- 225
 QY 200 KAGSFVASAPPAEMAVPOSSRLKYROLFNSHDTMSGHITGPQARTIMOSSLPQACL 259
 DB 226 ---TKAPLQGWVSPAKAKYDEIFLTKDKMDGFPISGLVAEILFKTGLPSTLL 277
 QY 260 ASINWLSIDIDQKILAEFTLAMHLIDVAM-SCQPLPPLVPEYIPPSFRVRSGSGMS 318
 DB 278 AHWSLCTKCGKLSKQFALAFHLISQKLIKIDPPHVLTPMIPPS--DRASLOKN 334
 QY 319 VISSSVQORLPEPSSDEQOPEKKLPVTFEDKKRENFEGSELKRRQALLQOORKE 378
 DB 335 ITGSSPV-----ADPSAIK-----ELDTLNEIVDLQREK 364
 QY 379 QERLAQERAEQERERERQOEAQRQLEKQERELERQERERERERERERAKR 438
 DB 365 ---NNVQDLKEK-----DITKQRTSVQ---DLQEVQEGENTNLQ 400
 QY 439 ELERQRLQEMRNRRQELINDRNKEQEGTVLKKARKTLFELBALDKKHQLEGKQDI 498
 DB 401 KLOAQKQ---QVQELLDL-----LDEQKQALEBQLEK 430
 QY 499 RCRLATQOEIESTNKSRELIAETHTLQOQLOESQOMLGRILPEKQILSDQLKOVQONS 558
 DB 431 RKKCAEEQLSSLK---AELTQESQISYEEELAK-----ARELSRLQOET 476
 QY 559 LHRDSLTLKALEKELARQOLREQLDEVERETRS---KLQELIDVFNNOIKELREIHSK 615
 DB 477 AE-----LEESVBSGKQLEPLQOHQDSQOEISSMQMKEMKEDLENHNSQLNWCSSP 530
 QY 616 QQL-----QKQSLAARLKQKQERKSLLEKQKQKDAQRVQERKQKQLEHVQ 665
 DB 531 HSILVNGATDYCSLSTSSSETANLNEHVQGSNLESEP-----THQES 573
 QY 666 EEOPRPRKHEEDRLKREDSVRKKEAEERAKPEMOD--KQSRLFHPHQEPAKLATQAPW 722
 DB 574 PSDFVGNPFGDPPKGS-----PFASDCFFROS-----TDPF 607
 QY 723 STTEGPIITIAQESVKVYVYRALYPFESRSHDELTIOPGDIWVWDSQT-----GEP 775
 DB 608 ATSTDPSSAANNSSITSV-----ETLKHNDPFAPAGGVVAASASATDPFSAVFQNE 659
 QY 776 GMLGQ-----ELKGTGWFPAN-----YAEKIENENVPYPAKAVTDLTS 814
 DB 660 SFGGGAADFTLSKVNNDPFRSATSSSVSNVITKVFETSVKSEDEPAPAL----- 713

QY 815 AAPKLAIRETPAPLPVTSSEP-STTPNNMADFSSTWSSSENEPETNDMTAAQPS 871
 DB 714 ---PKIGTPTRPCPLPQNDSPKEXPEMFCDPPTSATTTNNKEADSPNFANFSAYPS 768
 RESULT 6
 US-09-664-761-33620
 ; Sequence 33620, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmlica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; PRIOR FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 33620
 ; LENGTH: 90
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AP000193.1
 ; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.2
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2

OTHER INFORMATION: EST_HUMAN HIT: AA773823.1, EVALUATE 2.00e-37
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
US-09-864-761-33620

Query Match 6.8%; Score 429; DB 10; Length 90;
Best Local Similarity 94.0%; Pred. No. 1.4e-14;
Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 767 VDESQTGEPGWLGGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKLAIRETP 826
DB 1 VDESQTGEPGWLGGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKLAIRETP 60
QY 827 APLPVTSSPSTTPNNWADFSSTW 850
DB 61 APLAVTSSPSTTPNNWADFSSTW 84

RESULT 7

US-09-864-761-43004
Sequence 43004, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43004
LENGTH: 90
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
OTHER INFORMATION: EST_HUMAN HIT: AA773823.1, EVALUATE 2.00e-37
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
US-09-864-761-43004

Query Match 6.8%; Score 429; DB 10; Length 90;
Best Local Similarity 94.0%; Pred. No. 1.4e-14;
Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 767 VDESQTGEPGWLGGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKLAIRETP 826
DB 1 VDESQTGEPGWLGGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKLAIRETP 60
QY 827 APLPVTSSPSTTPNNWADFSSTW 850
DB 61 APLAVTSSPSTTPNNWADFSSTW 84

RESULT 8

US-09-864-761-33635
Sequence 33635, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33635
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000311.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 2.80e-01
; OTHER INFORMATION: EST_HUMAN HIT: AA815076.1, EVALUATE 1.00e-08
US-09-864-761-33635
```

```

Query Match          6.8%; Score 426; DB 10; Length 99;
Best Local Similarity 93.5%; Pred. No. 2.1e-14;
Matches 86; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 348 TFEKKRENERGSGVLEKRRQALFEQRRKEQRLAQLEBAERKREKREOEAKRQLE 407
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1 TFEKKRENERGSGVLEKRRQALFEQRRKEQRLAQLEBAERKREKREOEAKRQLE 60
```

```

QY 408 LEKQLEKQRELEERQEEERREKREIERREAKRE 439
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 61 LEKQLEKQRELEERQEEERREKREIERREAKRE 92
```

```

RESULT 9
US-09-864-761-46565
; Sequence 46565, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46565
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000117.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 9.00e-45
; OTHER INFORMATION: EST_HUMAN HIT: AA73823.1, EVALUATE 2.00e-37
US-09-864-761-46565
```

```

Query Match          6.7%; Score 422; DB 10; Length 82;
Best Local Similarity 95.1%; Pred. No. 2.7e-14;
Matches 78; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

QY 767 VDESOTGEPGWIIGGELKKTGTGFANPYAEKTPENEVPPRAKPYVDLTSAPPKLALRETP 826
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1 VDESOTGEPGWIIGGELKKTGTGFANPYAEKTPENEVPPRAKPYVDLTSAPPKLALRETP 60
```

```

QY 827 APLVTSSEPTTPNNMADFS 848
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 61 APLVTSSEPTTPNNMADFS 82
```

```

RESULT 10
US-09-864-761-34169
; Sequence 34169, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```

PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 34169
 LENGTH: 87
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000049.1
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
 OTHER INFORMATION: EST HUMAN HIT: AA815076.1, EVALUATE 4.00e-09
 OTHER INFORMATION: SWISSPROT HIT: Q35601, EVALUATE 1.80e-01
 US-09-664-761-34169

Query Match 6.6%; Score 413; DB 10; Length 87;
 Best Local Similarity 96.6%; Pred. No. 7.9e-14;
 Matches 84; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 348 TFFDKKRENERGSEVLEKRRQALLLEQKREOERLQALERAERKREKREOEAKRQLE 407
 Db 1 TFFDKKRENERGSEVLEKRRQALLLEQKREOERLQALERAERKREKREOEAKRQLE 60
 QY 408 LEKQLEKORLEKREERERKEIERRE 434
 Db 61 LEKQLEKORLEKREERERKEIERRE 87

RESULT 11
 US-10-144-621-2
 Sequence 2, Application US/10144621
 Patent No. US20020172671A1
 GENERAL INFORMATION:
 APPLICANT: Hsu, Yen-Ming
 TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street

CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/144,621
 FILING DATE: 13-May-2002
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/671,354
 FILING DATE: 27-JUN-1996
 APPLICATION NUMBER: US 08/486,344
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: US 08/484,709
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: US 08/475,894
 FILING DATE: 07-JAN-1995
 APPLICATION NUMBER: US 08/475,710
 FILING DATE: 07-JAN-1995
 APPLICATION NUMBER: US 08/474,697
 FILING DATE: 07-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Myers, Louis
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: 10274/009005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 659 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 FEATURE:
 LOCATION: 29...29
 OTHER INFORMATION: where Xaa at position 29 is Arg, Cys, Gly, or Ser
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-10-144-621-2

Query Match 5.2%; Score 325; DB 9; Length 659;
 Best Local Similarity 27.0%; Pred. No. 1.2e-08;
 Matches 91; Conservative 67; Mismatches 115; Indels 64; Gaps 15;

QY 914 LYPWAKKDNHINPKSVITVLEQD-MWFFGVQGGKMFPSYKLTISGPARKSTSI 972
 Db 1 LYPVLRHDDXETISVGIITNIRKEDGXWEGQINGRGLFPNFAREIKKEKKOPLT 60
 QY 973 DTGPTEPSASIKRYVSPAPKPAIPGEFEI-----AMTYESESQGLDFQO 1018
 Db 61 NKAP-EKP--LHEV--PGCNLLSSETLTITNKGRRRRRRCQVAFSYLPONDELKLV 115
 QY 1019 GDVI-VVTKKGDWMTGTGVGKSGVFPNSVYR-----LKDSGS 1056
 Db 116 GDIIIEVGEVEEGWEGVINGKTGFPNSFIKELSGSDDELGISQDBQLSKSLRETTGS 175
 QY 1057 GYAGKGTSLGKKPELQAVIAYATGPOL-----TLAPQQLLIRKKNGGWBEELQ 1110
 Db 176 ESDGSDS-STKSGANGTAVTAIIPKRVGVGFQDI PKDKPIKLPRRS---IEVEND 230
 QY 1111 ARGKKRQIG-WFPANYVLLSPGTSKITPTLPTAVQPAVCQYIGWYDTAQNDELAF 1169
 Db 231 FLVPEKTIKGLKLPAT---TATPDSK---TEMDSRTSKYCKYI--FPYEAQNDDELTI 282
 QY 1170 SKGQILINLAKN--DPDWKGEVSGQVGLFPSNYKL 1204

Db 283 KEGDIVTLINKCIDVGMWEGELNGRRGVFPDNFVKL 319

RESULT 12

US-09-879-957-133

Sequence 133, Application US/09879957

Patent No. US20020034755A1

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

HOFFMAN, No. US20020034755A1h

KAY, Brian K.

FOWLES, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 133:

SEQUENCE CHARACTERISTICS:

LENGTH: 59 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 133:

US-09-879-957-133

Query Match 5.1%; Score 322; DB 10; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 KVVVYRALYPPESRSHDEITIQPGDIVVWDSQTEGPGWLGELKGTGFWFPANYAEKI 797

Db 1 KVVVYRALYPPESRSHDEITIQPGDIVVWDSQTEGPGWLGELKGTGFWFPANYAEKI 59

RESULT 13

US-09-764-868-744

Sequence 744, Application US/09764868

Patent No. US2002016871A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT232

CURRENT APPLICATION NUMBER: US/09/764,868

CURRENT FILING DATE: 2001-01-17

Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 1510

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 744

LENGTH: 287

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (146)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (247)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-868-744

Query Match 5.0%; Score 316.5; DB 9; Length 287;

Best Local Similarity 28.8%; Pred. No. 1.3e-08;

Matches 88; Conservative 54; Mismatches 115; Indels 49; Gaps 11;

QY 919 AKKDNHNFNSDVIITVLEQDM-WMFGEVQOGKGMPPKSYVKLISCPVAKSTSIDTGP 977

Db 5 AEQVDILNLTSGEIVLYLEKIDTWGRNCRNOIGIFPANYVKI-----IDI-PE 54

QY 978 ESPASIKRVASPAKPAIPGEFIAMWTYSSBEGDITPOOGDVIYVTK-KDGDWMTGTV 1036

Db 55 GNGKRRECVSSHCKV---GRCVAREYIGEQKDELFSBEGEIIILKEVYNEEMARGEV 110

QY 1037 GDKSGVPPSNYVR-----LKDSESGTAGKTGSLGKKPEIAQVIAS 1077

Db 111 RGRITPPLNFVEPEVEDYPTSGANVLSTKVPLKTKXEDSSNSQVNL--PAEWCEALHS 168

QY 1078 YAATGEQUTLAPGOLILIRKANGGMEGELQARKKROIGWFPANYVVKLLSGTSGIT 1137

Db 169 FTAETSDLSFXRBDRIQILERLDSWCRGLQPRE-----GIFPAVFR-PCPAEKSM 222

QY 1138 PTELKPTAVPAGVQVIGMYDYTAONDELAFSGQIIVNLKEDPMMWGEVSGQVGLF 1197

Db 223 LAIVPKRKAQA-----LYDFRGENDELXFKAGXIITELBSVDDMMGELMGKSGIF 276

QY 1198 PSNYVK 1203

Db 277 PKNYIQ 262

RESULT 14

US-09-879-957-135

Sequence 135, Application US/09879957

Patent No. US20020034755A1

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

HOFFMAN, No. US20020034755A1h

KAY, Brian K.

FOWLES, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 135:
SEQUENCE CHARACTERISTICS:
LENGTH: 62 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 135:
US-09-879-957-135

Query Match 5.0%; Score 314; DB 10; Length 62;
Best Local Similarity 98.3%; Pred. No. 3.2e-09;
Matches 59; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1070 EIAQVIASVATGPEQLTAPGQILIRKKPGWMEGELQARGKKRGWFPFANYKLL 1129
Db 3 EIAQVIASVATGPEQLTAPGQILIRKKPGWMEGELQARGKKRGWFPFANYKLL 62

RESULT 15
US-09-879-957-134
Sequence 134: Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
McCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 56 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 134:
US-09-879-957-134

Query Match 4.9%; Score 309; DB 10; Length 56;
Best Local Similarity 98.2%; Pred. No. 4.9e-09;
Matches 55; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 907 EGLQALYPMRAKKDNLNFKSDVITVLEQDDMMWFGEVQGGKGFPPKSYVXLI 962
Db 1 EGLQALYPMRAKKDNLNFKNDVITVLEQDDMMWFGEVQGGKGFPPKSYVXLI 56

Search completed: March 14, 2003, 12:17:46
Job time : 479 secs



GenCore version 5.1.4 ps 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:05:51 / Search time 25 Seconds
(without alignments)
4664.442 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFTPTFGSGLDVMAITVE.....VGLFPSNVYKLTDDMPDPSQ 1213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :
1: p1r73:*
2: p1r12:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5063.5	80.8	1270	2 T09194	adaptor protein in
2	1735	27.7	1011	2 T13055	dynam associated
3	1698.5	27.1	1084	2 T13053	dynam associated
4	1333	21.3	1097	2 T31504	hypothetical prote
5	506.5	8.1	751	2 T34490	hypothetical prote
6	486	7.8	897	2 A54696	EGF receptor subst
7	485.5	7.7	896	2 S43074	epidermal growth f
8	410.5	6.5	1480	2 S48440	poly(A)-specific r
9	403	6.4	1381	1 S45781	probable calcium-b
10	391.5	6.2	1407	1 S28589	trichovyalin - rab
11	383.5	6.1	1909	2 A45592	liver stage antige
12	381	6.1	1898	1 A45973	trichovyalin - hum
13	375.5	6.0	1549	1 A40691	trichovyalin - she
14	359.5	5.7	737	2 T15597	hypothetical prote
15	359.5	5.7	1017	2 T15598	hypothetical prote
16	342	5.5	2526	2 T20531	hypothetical prote
17	342	5.5	2722	2 T20532	hypothetical prote
18	340	5.4	2738	2 E88320	protein F07A11.6 f
19	331.5	5.3	1069	2 T00377	KIA0642 protein -
20	330	5.3	793	1 JH0628	caldesmon - human
21	327	5.2	1794	2 T38459	hypothetical diver
22	324.5	5.2	1181	2 C86349	F8K7.4 protein - A
23	323.5	5.2	408	2 T42650	hypothetical prote
24	321	5.1	665	2 JC7191	85K c-Cbl-interact
25	319.5	5.1	1027	2 T46481	hypothetical prote
26	318	5.1	699	2 E84565	hypothetical prote
27	316.5	5.0	816	2 T17257	hypothetical prote
28	314.5	5.0	887	2 G88484	protein F23F12.8 f
29	309	4.9	4574	2 G02520	plectin - human

30	308.5	4.9	1701	2 T09127	probable erythrocy
31	307.5	4.9	1110	2 I51116	NF-180 - sea lamp
32	306.5	4.9	771	1 A33430	h-caldesmon - chic
33	304.5	4.9	585	1 A24168	involucrin - human
34	304.5	4.9	958	2 S47179	hypothetical prote
35	302	4.8	1052	1 A44937	kinetoplast-associ
36	302	4.8	4684	2 A59404	plectin (imported)
37	299	4.8	1233	2 T30989	sarcone/threonine p
38	297	4.7	2101	2 A42184	nuclear mitotic ap
39	296.5	4.7	3488	2 T34418	hypothetical prote
40	296.5	4.7	4957	2 T03455	ALK protein - huma
41	296.5	4.7	5262	2 T03454	ALK protein - huma
42	296	4.7	4687	1 A39638	plectin - huma
43	295	4.7	729	2 T50989	hypothetical prote
44	294	4.7	746	2 T47237	myosin II heavy ch
45	293	4.7	2442	2 T08621	centrosome associa

ALIGNMENTS

RESULT 1

T09194
adaptor protein interseectin - African clawed frog
C.Species: Xenopus laevis (African clawed frog)
C.Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C.Accession: T09194
R.Yamashita, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castragnoli, L.; Cesareni, U. Biol. Chem. 273, 31401-31407, 1998
A.Title: Intersectin, a novel adaptor protein with two eps15 homology and five src homol
A.Reference number: Z16605; MUID:99030416; PMID:9813051
A.Accession: T09194
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1270 <YAM>
A.Cross-references: EMBL:AF032118; NID:g2642624; PIDN:AACT3068.1; PID:g2642625
A.Experimental source: cell type oocyte
C.Function:
A.Description: involved in endocytosis
C.Keywords: endocytosis

Query Match	80.8%	Score 5063.5	DB 2	Length 1270
Best local similarity	80.6%	Pred. No. 2.6e-198		
Matches 980	Conservative 98	Mismatches 123	Indels 15	Gaps 8
QY	1	MAQFTPTFGSGLDVMAITVEERAKHDQFLSKPIAGITTDQARNFFQSGLPQVLAQ	60	
DB	1	MAQFTPTFGSGLDVMAITVEERAKHDQFLSKPIAGITTDQARNFFQSGLPQVLAQ	60	
QY	61	IWALADNMNDGMDQVFSIMAKLIKLGQYQPLSTLPVWKKQPVVAISA--PARGIG	118	
DB	61	IWALADNMNDGMDQVFSIMAKLIKLGQYQPLSTLPVWKKQPVVAISA--PARGIG	120	
QY	119	GIASMPPLTAAVAVPMGSIPIVVGMSPLVSSVPPAAVPLANGAPVYIOPLPAFAHPAAT	178	
DB	121	GIVGIPPLAAVAVPMGSIPIVVGMSPLVSSV--TVPLISNGAPAVIQSHPAFAH-SAT	177	
QY	179	WPKSSFSRSGSGSLNKLQKAGSFVVASAPPAEVAWVPOSSSLKTRQLFNSHDKTMGS	238	
DB	178	LPKSSFSRSGSGSLNKLQKAGSFVVASAPPAEVAWVPOSSSLKTRQLFNSHDKTMGS	237	
QY	239	HLTGQARTIMQSSSLPQALASIMWLSLDIDQDGLTAEEITLMMHLIDVMSQOPLPV	298	
DB	238	NLTGQARTIMQSSSLPQALASIMWLSLDIDQDGLTAEEITLMMHLIDVMSQOPLPV	297	
QY	299	LPPEYIPPSFRRVSGSGMSVSSSVVDORLPPEPSSSEDOQPEKKLPVTEDEKKRNF	358	
DB	298	LPPEYIPPSFRRVSGSGSLSSSVVDORLPPEPSSSEDOQPEKKLPVTEDEKKRNF	357	
QY	359	RGSVLEKRRQALLQCKKEOERLAOLERAQEKKEEROQEKKROULEKQEKOREL	418	
DB	358	RGSVLEKRRQALLQCKKEOERLAOLERAQEKKEEROQEKKROULEKQEKOREL	417	

QY 419 EROEEERREKIERREAKKELEROROLEMERNRROELLNORNKEOEGTVLAKRKLTLE 478
 DB 418 EROEEERREKIERREAKKELEROROLEMERNRROELLNORNKEOEGTVLAKRKLTLE 477
 QY 479 FELBALNDKKGHLEGLKODIRCRLATROEIESTNKSREIRIAETIHL00OLOESQOMLG 558
 DB 478 FELBALNDKKGHLEGLKODIRCRLATROEIESTNKSREIRIAETIHL00OLOESQOMLG 537
 QY 539 RLIREKQILSDQKOVONSLHRDLSLTLKRALEAKELAQOOREOLDEVERERTSLOE 598
 DB 538 KMLPEKSLIDQKOVONSLHRDLSLTLKRALEAKELAQOOREOLDEVERERTSLOE 597
 QY 599 IDVENNOLKELREIHSKQOLOKORSLERARLKOEKOEKSLERKEDAKORVBERKO 658
 DB 598 IDVENNOLKELREIHSKQOLOKORSLERARLKOEKOEKSLERKEDAKORVBERKO 657
 QY 659 WLEHVQOEBOEPKPHBEDRLKREDSVRKKEAEERAKPEMOKQOSLPHPHOEPALAT 718
 DB 658 WQDVKOEER--RYKFODEEKEKESVQKCEVER--KPELOEKPNKPFHQPPPGKLG 713
 QY 719 QAPSTTEKGLTISAQSVKVVYRALYPESHDEITIOPGDIYMWDESQTEPGL 778
 DB 714 QIPMNTKAPLITIN-QGDVAVVYRALYPRDASHDEITIEPGDIYMWDESQTEPGL 772
 QY 779 GGBLKGTGWFPANYAEKIPENEVPTPAKVTDLTSAPAPKALRETPAPL-PVTSSEPS 837
 DB 773 GGBLKGTGWFPANYAEKIPENEVPTPAKVTDLTSAPAPKALRETPAPL-PVTSSEPS 828
 QY 838 TTPNNMADFSSTWSSSEKRETDNMDTMAOPSLTVPSAGOLQORSAFTPATGSSPS 897
 DB 829 TNSNMADFSSTWSSSEKRETDNMDTMAOPSLTVPSAGOLQORSAFTPATGSSPS 888
 QY 898 PVLGQGEKVEGLQOALYPMRAKKNHNFNSKSDVITVLEQODMWFGEVQOGKMPKPS 957
 DB 889 PVLGQGEKVEGLQOALYPMRAKKNHNFNSKSDVITVLEQODMWFGEVQOGKMPKPS 948
 QY 958 YVKLISGVNRKSTIDTGPTESSPAKRVASPAKPAIPGEFTIAMTYESSBEGDITFQ 1017
 DB 949 YVKLISGVNRKSTIDTGPTESSPAKRVASPAKPAIPGEFTIAMTYESSBEGDITFQ 1008
 QY 1018 QGDVIVVTKKQDMMWTVGDSKSGVFPNNYVRLKDESGTAGTGLSGKKPEIAQVIAS 1077
 DB 1009 QGDVIVVTKKQDMMWTVGDSKSGVFPNNYVRLKDESGTAGTGLSGKKPEIAQVIAS 1068
 QY 1078 YAAATPEQTLTAPGQILIRKKNPGMWEGELQARKKRQIGMFPANYVLLSGTISKIT 1137
 DB 1069 YAAATPEQTLTAPGQILIRKKNPGMWEGELQARKKRQIGMFPANYVLLSGTISKIT 1128
 QY 1138 PTELPKTAVQAVQVIGMVDYTAQNDELAFPSKQIINVINKEDPPMWKGEVSGVGLF 1197
 DB 1129 PTELPKTAVQAVQVIGMVDYTAQNDELAFPSKQIINVINKEDPPMWKGEVSGVGLF 1188
 QY 1198 PSNVYKLTITDMDPSQO 1213
 DB 1189 PSNVYKLTITDMDPSQO 1204

RESULT 2
 T13055
 dynamitin associated protein isoform 160-2 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13055
 R:Roos, U.; Kelly, R.B.
 J. Biol. Chem. 273, 19108-19119, 1998
 A:Title: Dap160, a neural-specific Epsins homology and multiple SH3 domain-containing pro
 A:Reference number: Z17594; MUID:98334647; PMID:9668096
 A:Accession: T13055
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-1011 <ROO>
 A:Cross-references: EMBL:AF054612; NID:g2996029; PID:g2996030; PIDN:AAC39139.1
 C:Genetics:

A:Gene: Dap160
 A:Cross-references: FlyBase:FBgn0023388

Query Match 27.7%; Score 1735; DB 2; Length 1011;
 Best Local Similarity 33.4%; Pred. No 2.4e-63;
 Matches 413; Conservative 202; Mismatches 332; Indels 288; Gaps 31;

QY 11 SLDWAITVEERAKHIDQOFLSKPIAGFITGDAARNFFQSGLPQVLAQIWMALADNMD 70
 DB 4 AVDAVAITPRERLKYQEQFRALQFOAGFVTGAQAKGFELQOLPPLILGQIWMALADNMD 63
 QY 71 GMDVPEFSIMAKLIKTLQGVOLPSTLPYPMKQOPVAISAPAFGIGIASMP---LT 127
 DB 64 GMDVPEFSIMAKLIKTLQGVOLPSTLPYPMKQOPVAISAPAFGIGIASMP---LT 113
 QY 128 AVAPVPMGSIPIVGMSPPLVSVPPAAVPPANGAPVYIOPLPAPFAHPATWPKSSFSR 187
 DB 114 SLSPLD-----PLKGI-VAVAAPVVPVAPVA--VAATVSP-PCVSVSPSGTPTSN--- 162
 QY 188 SGPSQNLTKLQKQSFVVASAPPAEAVVPOSSRLKROLFNSHDKTMGHLTGPOART 247
 DB 163 --PSRHTSISERAPSISVN---QGEWAQAAQRKXTYQVFANNDRTSSGLTGSQARG 217
 QY 248 IIMQSLPQAOALASTIMNLSDIIDQGLKTAEEFIIAMHLIDVAMSGOPLPVLPEYIIPS 307
 DB 218 VLVQSKLPQVTLAQIWTLSIDIDGRLNCDEFILAMFLCEAMAGEKIPVTLPOBWVPPN 277
 QY 308 FRRVSGSGMSVYSSSSVDORLPBESSD-----EQPEKLP-VTFEDK 353
 DB 278 LRKIRSRP---SVSGVSPRGSQPSARSHASVSQSGVGAVDADPAGLPQGTSPEDKR 333
 QY 354 RKNFERGSELEKRGQALLLEQORKEERLAQLEAEORKEERKEOEAKRQLEKOLE 413
 DB 334 KENVYQAELEDRRKRIEDQORKEERKEERKEERKEERKEERKEERKEERKEERKEER 393
 QY 414 KORELEOREEERREKIERREAKKELEROROLEMERNRROELLNORNKEOEGTVLAKR 473
 DB 394 KORELEMEKEBERKELAKAEAREKELKOEQOEKARIAEMNAQKREGERVLQKQAH 453
 QY 474 RTLEFELEALNDKKGHLEGLKODIRCRLATROEIESTNKSREIRIAETIHL00OLOES 533
 DB 454 NTQNLVELSTLNEKIKELISORICTRAGVTVKTVIDMRTQRTDSSEMSQLARIEQ 513
 QY 534 QQMLGRPLPEK-----QILSDQKOVONSLHRDLSLTLKRALEAKELARQOLR 582
 DB 514 NAKLQILTERAKMWAASKASGALGENAQEO-----LMAFAKQILINDIK 563
 QY 583 EOLDEVERETRSKLOEIDVFNQKLERE---IHSKQOLOKORSLEARLKOEKOEK 637
 DB 564 DKVENISKEISKEDINTNDVQMSLEKALSLITTKCEDLYKEYDVQRTSVLELKYNRK 623
 QY 638 SLELEKQKDAQRQVBERKQVLEHVQOEBOEPKPHBEDRLKREDSVRKKEAEERAKP 697
 DB 624 NETSVSSAWDT-----GSSSAW---BETGTTVDPY-----AVASNDISALAAP 664
 QY 698 EMQDKQSRLEFHPQEPKATQAPWSTTEKGLTISAQSVKVVYRALYPESHSHDEI 757
 DB 665 AYD-----LGGPAP-----EG-----FVKQAVVEFNANNAEIEI 693
 QY 758 TIOPGDIYVWDSQTEPGLGGLKGTGMFPANYAEKIPENEVPTPAKPVTLTISA 817
 DB 694 TFPVGDIIILVPLEQNAEPQWLAGEINGHGMFPESYVLEKLEGEV-----A 739
 QY 818 PKLALRETPAPLPVTSSESTTPNNWADPSSSTWSSSEKRETDNMDTMAOPSLTVPSA 877
 DB 740 PVAAV-BAVDQVADT-----YNDKIN-----SSIPAA 768
 QY 878 GQLRQSAFTPATATGSSPSVPLGQGEKVEGLQOALYPMRAKKNHNFNSKSDVITVLE 937
 DB 769 S-----ADLTAA-----DV----- 778
 QY 938 QQDMMWFGEVQOGKGMFPKSYVKLISGVRKSTIDTGPTESSPAKRVASPAKPAIPG 997

Db	779	-----	778
Qy	998	EFTAMTYEESBQDGLTFQOGDVI VTKKDGMMTGTGVGDKSGVPSNRYRLKDSBSSG	1057
Db	779	EYIAAYPEESAEBDDLFSAGEMWVIKKSGEMWTGTIGSTGMFPNSNYQKADVGTAS	838
Qy	1058	TAG-----KTGSLGKKPELIAQVIASYAATGPEDLTAPGLILIRKKNCGMWEGELQ	1110
Db	839	TAAEPVESLDQGMKAKSEIIAQVIAPYEATSTBEDSLTROLINIRKKTSGMWEGSLQ	898
Qy	1111	ARGKKRQIGMFPANVYKLLSPG--TSKITPTELPTAVQAPAVC-OVIGMYDYTAQNDEL	1167
Db	899	AKGRRRQIGMFPATVVKVLOAGRNSGRNTPVSGSRILEMTEQLIDKVIALLYPKAQNDEL	958
Qy	1168	AFSGQIINVLTNKEDPDMWKGVSGVGLFPSNRY 1202	
Db	959	SFDKDDIIISVLRGRDEPEWWRGELNLSGLFPSNRY 993	

```

RESULT 3
T13053
dynamlin associated protein isoform Dap160-1 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13053
R:Roos, J.; Kelly, R.B.
J. Biol. Chem. 273, 19108-19119, 1998
A>Title: Dap160, a neutral-specific Eps15 homology and multiple SH3 domain-containing protein
A:Reference number: 217594; MUID:98334647; PMID:9668096
A:Accession: T13053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1094 <RCO>
A:Cross-references: EMBL:AF053957; NID:g2984714; PID:g2984715; PIDN:AAC39138.1
C:Genetics:
A:Gene: Dap160
A:Cross-references: FlyBase:FBgn0023388

```

Query Match	27.1%;	Score 1698.5;	DB 2;	Length 1094;
Best Local Similarity	31.6%;	Pred. No. 8.1e-62;		
Matches 417;	Conservative 201;	Mismatches 329;	Indels 371;	Gaps 33;

```

Oy 11 SLUDWAIIVEEAKHDQOFLSLKPIAGFTTGQANFFQOSGLPQVLAQIUALDMND 70
Db 4 AVDMAAVIPREHLKTKQOEQFRALQPOAGFTGAQAGFFLQSLPPLLIQIUALDLTSD 63
Oy 71 GRMDOVEFSIAMKLIKLIKQYOLPSTLPVWKQOPVAISSAPAFGIGIASMP--LT 127
Db 64 GKMNINEFSIAKCLINLKLRGMDVPEKVLPPSL-----LSSL-----TGVPEPMTRGSTS 113
Oy 128 AVAPVPMGSIPIVWGMSPLVSVPPAAPVPLNGAPVPIQPIPAFPAHPATMPKSSFSR 187
Db 114 SUSPLD----PLKGI-VPAVAPVVPVAVPVA--VATVISP-PGVSVS8GPIPPPSN--- 162
Oy 188 SGPGSLNLTKLQKQASFDVASAPPAEMAHPQSSRLKYROLFNSHDKTMGHLTGPOART 247
Db 163 --PSPRHTSISBRASIESVN--QGEMAVQAQKRYQTVENANDRPSGYLTGSGQARG 217
Oy 248 ILMOSLIQAOCLASITWNLSIDIDQDKLTAEFTILMHLIDVAMSGOPLPVLPPEYIPPS 307
Db 218 VLVOSKLTPOVTLAQITWLTSDIDGDRLNCDEFILAMFLCEKMAAEKIPVTLPOEWMPN 277
Oy 308 FRVRSGSGMSVISSSSVQORLPEPPSED-----EQOPEKPLP--VFEDFK 353
Db 278 LRKISRRC----SYSGVVRPGSQPASHASVSSQSGVGVVDADPTALPQOTSFEDKR 333
Oy 354 RENFERGVSYLEKRRQALIEQQRKEOERLAQLEPAEOEKEREROEOAKROLELEKOLE 413
Db 334 KENYVAGQALDRRRKIMEDQOKEREEREREREADREKARLEAEKQOEBELERQLO 393
Oy 414 KORELEROEEERREKREIEREAKRELEOROLEWERNRROELTYORNNEOGGTVLKAR 473
Db 394 RORELEMEEBORKELEKAPKRELEKQOQEWQOARIENNAQKEEEOERVLKQKH 453

```

QY	474	RKTLPEFLNKNKQKQLEGGKQLODIJRCRLATOROEIESTNKSRELRJAEITHLOOQOE	533
Db	454	NTQLANVELSTLNKRIKELSORICDTRAGATNVKTYIDGMRFORDTSMSEMOQLARIMEQ	513
QY	534	QOMLGRILPEK-----QIISDQKVOYONSJLHRDLSLTLKRLAEKELARQOLR	582
Db	514	NAKLQLOTORAKWEAKSKASGALLGGENAOEO-----LNAFPAHKOLINDIK	563
QY	583	EOLDEVERETRSKLOEIDVFNNOLEKRE-----JHSKOOLOKRSLAEARKKQKQERK	637
Db	564	DKENISKELIESKEDDINTNDVOMSLTKELSLALITTCEDLYKEIDVORTSVLEKTKYRK	623
QY	638	SLELEKQKEDAOARVOERDKOMLEHVQOEOPRPKPHBEDRLKREDSVRKKEAEERAKP	697
Db	624	NETSVSSAMDT-----GSSGAM-----EETGTTVDPY-----AVASNDISALAAP	664
QY	698	EMQDKOSRLPHPHOEPAKLATQAPMASTTEKGPLTISAGESVNVVYRYRALYPRESSSHEI	757
Db	665	AVD-----JOGPAP-----EG-----FVYQAOVYEFHNAKMAEI	693
QY	758	TIOPGDIWMTDESQTEGPBGLGELKGTGKWFEPANYAEKIPENEVPTPAKPVTLDTLSAPA	817
Db	694	TFVPGDIIIVLPLEQNAPEGLAGEINGHGWPESSVEKLEBGEV-----A	739
QY	818	PKLARETPAPLPEVTSSEPTTPNNMADFSSTPSSSNKEKPETDNMDTMAOPSLTPVSA	877
Db	740	PVAAY-BAIPDAQVADT-----YNDINT-----SSIPAA	768
QY	878	QQLRQBSAFPPAPATATSSSPPEVLGQEKVEGLOAQALYWRKAKKONHLPNKSVDITVLE	937
Db	769	S-----ADLTAAQ-----DV-----	778
QY	938	QODMMWGEVQOGKGFPKSKYVLISGPVRKSTSIDTGPTEPSAGLKEVAPAAKPAIPG	997
Db	779	-----	778
QY	998	EEPIAMTYTSSSEOGDITFPOQGVIVYTKKQDWMVTGYDKSGVFPSPNYRYRLKO-----	1052
Db	779	EYVIAAYPYESAEBGLDSFAGEMVWVVIKKEBEMWTGITIGSRMGFPSPNYVOKADVGTAS	838
QY	1053	-----SEGG-----	1057
Db	839	TAAAEPEVESLDQETTLNGNAATYAATVAEVOBOVYQPLPVQESSEOPISPGVGAEEAHED	898
QY	1058	-----TAGKT-----GSLQKKEEIAQVIAASYATGPEOLT	1087
Db	899	LDTEVQSINTQSTQOSEPAESYSRMSRTSSMTPGMRAKSEIAQVIAPEATSTEQLS	958
QY	1088	LAPQOLLIRKKNGGWBEELQARKKQOIGMFPFANYKYLSPG--TSKITPTPLPRTA	1145
Db	959	LTRQOLLIMIRKKTDSGWBEELQAKRRROQIGFPAITYKYVLQGRNSGRNTPVSGSRTE	1018
QY	1146	VQPAVC-QVIGMVDYTYQONDDLAFESKQOINVNLNDEDDPMKGEVSGOVGLFPSPNY	1202
Db	1019	MTBQIDLKVIALYPRKQNDDELSPKXDIILISVLRGDEPMWRGELNGLSGLFPSPNY	1076

RESULT 4
T31504
hypothetical protein Y116A8C.36 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31504
R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21041
A:Accession: T31504
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1097 <MIL>
A:Cross-references: EMBL:AL117204; PIDN:CB55138.1; CESP:Y116A8C.36
A:Experimental source: clone Y116A8C

C:Genetics:
A:Gene: CESP.Y116A8C.36
A:Introns: 50/3; 245/3; 411/3; 486/3; 697/1; 820/3; 926/2

Query Match 21.3%; Score 1333; DB 2; Length 1097;
Best Local Similarity 29.4%; Pred. No. 5.5e-47;
Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;

QY 15 WAITVEERAKHDQOFLSKPIAGFITDQ-----ARNFFQSGLPQVLAQIWMALADM 67
DB WEVSFAEYOKNPAMF-----GOLTGQPPMDAVTARNALMRNSLPTQVLSQIWMALSDL 57
QY 68 NNDGMDQVRSIANKLIKLGQYLPSTLPVVKQOPVAISSAPAFGIGISMPELPT 127
DB 58 DKDGLDREYSIAWRLANCLAGIPPPQLPSLLVPA--RNAAPTWS----- 107
QY 128 AVAPYPMGSIIVVGMSPPLV-----SSVPAVPLANGAPVYQPLPAFAHPAATWPK 181
DB 108 -----RHGSVYQSOTLPALIDRRMSQSYIPSA--PVSIAQTP----- 142
QY 182 SSSFRSGPGSOLNTKLOKQSFVYASAPPAEWAVPOSSRLKTRQLPNSHDKTMSGHLT 241
DB 143 SSRHNSISAGSPLN-----NDRVFEGRQLENWAIPIHNNKLKYSOLFALDKERLGLSLS 196
QY 242 GPQARTIMQSSLPQALASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 301
DB 197 SUGRSALGSLGFLPNTVAHIMFLSDVAKDKLSVDEVAISQYIEMKSGFALPKITPL 256
QY 302 EYIPPSFRVRVSGSGMSVSSSVDOQLPEEPSEDEQOPEKKLPVTTEDKKRENERGS 361
DB 257 ELV-----RMCISRSRANPTPELEPGAEPPQ--KSPAPKTFEDKRODNLKSGQ 303
QY 362 VELEKRRQALLQOKKEGERLAQLEBAQERKERERQOEAKRQOLELEKQERLEERQ 421
DB 304 ALELERRVLEEBEORRAVEKEREKEBEAKKNERQOEKQAEVEKQAELEKORITIEAQ 363
QY 422 REEE--RRKEIER--BAKKELE--RQOLEMENRRQOELNCKNKQOEGTIV 469
DB 364 REEBEKKRRLMERREDEKMKRKQVMEKAKVQKQWPNQKQWPNQKQENELAQ 423
QY 470 LKARKKTLFELEBALNDKQKQLEGLQDIRCLATQROEISTKSKRELIATITHLQQO 529
DB 424 RQOKRKTQLOFQALDEKVIDEVDIGAKAEVAEVTGFIKMRSTRDEKVARI--KE 479
QY 530 LOESQOMGRLLPEKQISDOLKQVQNSLHRDLSLTLKALKEKELARQOLREOLD-- 586
DB 480 LOETNQ--KTAIESQELGHQ--LQKOSAKET--TQKSLBELAKRKRDARIRAIEDIA 533
QY 587 -EVERETSKLOEIDVF--NNQLEIREIHSKQOLQKORSLAARLKQOEKRSLELEKQ 644
DB 534 LELSTEKEKSYNQTEIILKTNKEKYKTQVSK-----LVAKREERYNSFEL-- 578
QY 645 KEDQORRQOERDKQWLEHVQOEOEOPRRKP-----HEDPLKEDSDVRK 688
DB 579 ----LVHAQTARSKIGFEPEKASAPASAPAPAPAPATTNGPANNNDAFGEFD--K 631
QY 689 KEAERAPKEMQDKOSRLFHPHOEPAKLATQAPWSTTEKPLTISAQESVYVVYRALYP 748
DB 632 TDAQRFADGATST-----ADPFAQIAQAP--ASHKGVDSQAFNIHDTYCRALLFA 683
QY 749 PESRSHDITITOPGDIWVWDSQTEGPEWLGELKGTGMFPYAYAEKIPNEVPTAKP 808
DB 684 FEAREDELESEPGVITVQSHAAPEGMAGOLREKVGWPEAFVEAIA--AVPTG-- 739
QY 809 VTDLTSAAPKALARETAPLPVTSSEPTTPNNWADFSSTWSSSNEKPEPTDWDWTAA 868
DB 740 -----GDPPIQMPNMTPTSSVDOIGV--KAARKAEI-----AA 772
QY 869 QPSLTVPAGQLRORSAFTPATGATGSSPVVLGGEXVEGLQALVPMWAKDNHNLNF 928
DB 773 AMGLTEGA-----PASSAPAAAVIS-----QCTAIOFQWRANREDISFA 814
QY 929 KSDVITVLEQDDMMWFG--EVQOGQKMPFKSYVKLISGVKRSSTSIDGPTSPASLAKVA 987

DB 815 KQDTLEVELEKQMKWKGRNPAGEICMFPKSYKEVGAATTSTTTP--VSPSASAGAPCA 873
QY 988 SPAAKPAIPGEE-----FIAMTYSESQODLTFQOQDVIVTKKQGDWMTG 1034
DB 874 AGAQDVVSDVTLQASETAPQOQLYTYIYPEANETTDLMHGDITLVEKNDKEMWK 933
QY 1035 TWGDKSGVFPNSYVRLKDSGSGTAGKT--GSLGKKPEI---AQVIASATATGPOLTLA 1089
DB 934 RCGREGIFPANYVEISVQOAGDPPTPQAPPAFPVLCBAKVVDVVASAPQGLIK 993
QY 1090 PGQLIRKKNPQMGWEGELQARGKKROIQMPFANYVVLSPGTSKITPTLEPKTAQVA 1149
DB 994 VGEIVKIRKSAAGWMEGELINGKP--IAGNFGSYVAVLEASPAI----- 1040
QY 1150 VQOVIGMYDTYANDDELAFSKQIINVLNKEDPDWMK--EVSQVGLFPNSYVK 1203
DB 1041 --RATVYDYEASQPDDELGFTGTGVITVTDKSEAEWMSGHREODPSKGLFPNSYVQ 1095

RESULT 5

T34490
hypothetical protein ZK1248.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
R:latreille, P.
Submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid ZK1248.
A:Reference number: Z21534
A:Accession: T34490
A:status: preliminary; translated from GR/EMBL/DBJ
A:molecule type: DNA
A:Residues: 1-751 <LAT>
A:Cross-References: EMBL:U29244, PIDN:AACT1084.1, GSPDB:GN00020, CESP:ZK1248.3
A:Experimental source: strain Bristol NZ, clone ZK1248
C:Genetics:
A:Gene: CESP:ZK1248.3
A:Map position: 2
A:introns: 37/1; 74/3; 118/3; 355/2; 661/1; 728/2

Query Match 8.1%; Score 506.5; DB 2; Length 751;
Best Local Similarity 24.7%; Pred. No. 1.2e-13;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;

QY 15 WAITVEERAKHDQOFLSKPIAGFITDQARNFFQSGLPQVLAQIWMALDMNNDGMD 74
DB 111 WPIPTDQAKYDSTIQSLNPNVNGKLSGAHVPRVLMNSGLDHALAKIWEISDQDKGNLD 170
QY 75 QVEFSIAMLKLIKLGQYLPSTLPP--VMKQOPVAISSAPAFGIGIASMPPLTAAVAV 132
DB 171 RIEMSVALLHVYRSLSQSDPVAQPLPNNLIHPSKAMVYAHSSPNF-----AAPP--HPR 221
QY 133 PM-----GSIPVYVGSPPVSSVPPAAVPLANGAPVYQPLPAFAHA----- 176
DB 222 PMWGRAGSVTSLDVIDNMSQ--YSATMPRAYP-----QPRASASAAHVNGSFTS 271
QY 177 -ATWPKSSFSRSGSGLNTKLOKQSFVYASAPPAEWAVPOSSRLKRYQLFNSHDKT 235
DB 272 GASTPISASH-----IHSF-----PAEWINTGD--YADQFAQIDTN 308
QY 236 MSGHLTGPOARTIMQSSLPQALASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGPL 295
DB 309 KQGLVDGMKRAPMWTTGISAQILAHVALADIKKCGQLEQFALTTHLLDMARKGESI 368
QY 296 PVLTPREYIPSPFRVRVSGSGMSVSSSVDOQLPEEPSEDEQOPEKKLPVTTEDKKRE 355
DB 369 PSELPLHLIPSPFRPTBSALHPAQSISTPQLPEAISMETKEALE-----GE 417
QY 356 NFERGSVELLEKRRQALLQOEKQERLAQLEBAQERKERERQOEAKRQOLELEKQLEKQ 415
DB 418 NEE--WKQLAESIQGMIVKRTAEAVIQLLE-ADMTITNGSSIK--NLQVELATLBESTV 470

QY 416 RELEROREBERK-----ETERRAKRELEROROLEMERNRROELLNORNEQEGTVV 469
 Db 471 KOLEROGKATRLADYDTQIEOLESA---CKAQKETKEDEKRMQOIDEADNAED--- 524
 QY 470 LKRRKTLFELEALNDK--HOLEGLDORLATOROIESTNKSRELRIAEITHQ 527
 Db 525 CKA-----NDEKEMELKEIEMLDNOFKTVRGEIVKETSOREQKAEILT--- 569
 QY 528 QOLOESQOMLGRLLPEKQILSDOLKQVONSILHRSLTLTKRALKEAKELARQOLR-EOLD 586
 Db 570 -----TLERRE-ARDQIQHEKID 586
 QY 587 EYERERKSLQELIDVENNOLKELREIHSKOQLOKORSLAARLKQKQERKSLIEKOK 645
 Db 587 AAIENTTKLEQVSDAVEKSEEMVQI---LRSQORLSTVIDSLDSDDTVYETAG 641
 QY 646 EDQORVQERDKQMLEHVQOEOBRPRKPHBEDRLKREDSVRKKKEERAKPEMQOKSR 705
 Db 642 TSSQ-----NHVQPPDD-----FASARNPAD----- 665
 QY 706 LPHHQBPAKLATQAPWSTTEKGLPTISAQESVYVYALYPEERSHDEITIQPDIV 765
 Db 666 -----PFAQYDQ-----FGSSGH----- 678
 QY 766 MWDESQTGEWGLGELKGTGMFPAN--YAEKI PENEVTPAKPVTLTSLAPAKLALR 823
 Db 679 -PFAAFPTDFPAQGG-----FPSPSGFAQSAAPAK---PAP-----RPAPPKSAR 719
 QY 824 ETPAPLPVTSS--PSTTPNNMADFS 847
 Db 720 ETPVNDPFAPOGOSTOPAGFADFA 744

RESULT 6

EGF receptor substrate eps15 - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-Nov-1999
 C/Accession: A54696
 R/Fazio, J. F.; Mutch, J. L.; Matoskova, B.; Wong, W. T.; Di Fiore, P. P.
 Mol. Cell. Biol. 13, 5814-5828, 1993
 A>Title: eps15, a novel tyrosine kinase substrate, exhibits transforming activity.
 A/Reference number: A54696; MUID:93361014; PMID:7689153
 A/Accession: A54696
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-897 <PAZ>
 A/Cross-references: GB:L21768; NID:9404756; PIDN:AAA02912.1; PID:9404757
 C/Superfamily: calmodulin repeat homology
 C/Keywords: EF hand; phosphoprotein
 F/48-80/Domain: calmodulin repeat homology <EF1>
 F/160-192/Domain: calmodulin repeat homology <EF2>
 F/223-255/Domain: calmodulin repeat homology <EF3>

Query Match 7.8%; Score 486; DB 2; Length 897;

Best Local Similarity 21.3%; Pred. No. 1e-12;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

QY 15 WATTEBRAGHQOFLSLKRIAGFITCDARNFQSGLPQVLAQITMALADNNNGRMD 74
 Db 122 WAKSBDKAKYDAIPBLSFVDFSLGDKVKPVLLNSKIPVELIGVWELSDIDHDGKLD 181
 QY 75 QVEFSIAMKLKILKLOGYQUPSTLPVMKQOQVAISSAPAFIGIGIASMPPLTAIVAPVM 134
 Db 182 RDEFAVAMFLVYCALE-----KEPVM 203
 QY 135 GSIPVGMSEPLVSVPPAAVPLANGAPVIOPLPAHPAATWPKSSSFSSGFGSQOL 194
 Db 204 -----SLPPALVPP-----SKR 215
 QY 195 NTKLQKQASDVASAPPAEMAVPQSGRLKYROLFNSHDKTMSGHLTGPOARTILMOSSL 254
 Db 216 KT-----WVSPAKKAKYDEIFLTKTDKMDGVSGLEVRFTPLTKGL 257

QY 255 PQOQASIMNISDIDODCKLTAEEFIAMHLIDVAM-SGQPLPVLPPEYIPSPFRVRS 313
 Db 258 PSALLAHIMSLCDTGCCKLSKMDQALAFHLINQKILKIGIDPPHSLTFEMIPSP---DRS 314
 QY 314 GSGMVISSSSDVQRLPREPSSDEQOPEKPLPTFPEKKRENERGVSLEKROALLE 373
 Db 315 SLQKNITGSSPV-----ADFSAIK-----ELDTINNEIYD 344
 QY 374 QORKEORLAQLEAERERERERQOEAKROLEKOLEKORELEROREBERREKIERR 433
 Db 345 LOERK-----NVEQDLKEKE-----DYVQKRTSEVQ--DLODEVORE 380
 QY 434 EAKKELEROROLEMERNRROELLNORNEQEGTVVLKRRKTLFELEALNDKQOLEG 493
 Db 381 SINQKQAOQK-----QVQELLGE-----LDQKQOLEE 410
 QY 494 KLQDRLRATOROIESTNKSRELRIAEITHLQOQLOESQOMLGRLLPEKQILSDOLKQ 553
 Db 411 QLOEVRKKCAEBAOLISSLK-----AEITSQSOISSYEEELIK-----AREELSR 456
 QY 554 VQONSILHRSLTLTKRALKEAKELARQOLEQDEVRE-----TRSKLOEIDVNNOLKE 608
 Db 457 LQOETAQ-----LEESVSGKAQLEPLQOHLQBSQOELISSMQRLEKQDLETDNNQ--- 507
 QY 609 LREIHSKOOLQK-----RSLAARLKQKQERKSLIEKOKEDARQVQ 653
 Db 508 -----SNWSSPQSVLVNGATDYCSLSTSSSETANFNHAGQNNLSEPHQSSSVRS 562
 QY 654 ER-----DKWLEHVQOEOBRPRKPHBEDRLKREDSVRKKA---BERAKPEMDK 702
 Db 563 PEIAPSDVTDESEAVTVAGNEKVTPR--FDQDKHKSKEDEPNVSSSLTDVADTNDLDF 620
 QY 703 GSRLEF-----HPHQBPAKLATQAPWSTTEKGLPTISAQESVYVYALYPEERSHDEITIQPDIV 743
 Db 621 QSDPVPVSGDPPKDPFGKIDPFGGDPFGKSDPPFASDCEFKQSTDPPTTSTDPFASSN 680
 QY 744 RALVFPESRSHDEITIQPDIVMWDESQTEP--GWLGLGELKGTGMFPANVYAEKIPENE 801
 Db 681 SNTSVETWIKNDPAPGCTVVAASDQATDPFASVFGNESFG--DGFADFTLTKANNED 739
 QY 802 VPTP-----AKVTVTLTSA-----PA--PKIALRETPAPLPVTSSESTTPNN 842
 Db 740 AENPTISSSTSSVTIAKMLEETRSKSEDPVAPLPKVGTPTRCPPP-----GKGRPIN 794
 QY 843 WADFSST-----WPSSNKEPEPDMDMTMAAPSLVPSAGQLRQSAFTPATATGSS 895
 Db 795 KLDSSDPLKLNDFPQPPGNDSPKEXPDWFC-----DPTSSITTNKE 838
 QY 896 PSP 898
 Db 839 ADP 841

RESULT 7

S43074
 epidermal growth factor receptor substrate - human
 C/Species: Homo sapiens (man)
 C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
 C/Accession: S43074; I38525
 R/Bernard, O. A.; Mauchautte, M.; Mecucci, C.; van den Berghe, H.; Berger, R.
 Oncogene 9, 1039-1045, 1994
 A>Title: A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4/
 A/Reference number: S43074; MUID:94181254; PMID:8134107
 A/Accession: S43074
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-896 <BER>
 A/Cross-references: EMBL:Z29064; NID:9470034; PIDN:CAA82305.1; PID:9470035
 R/Wong, W. T.; Kraus, M. B.; Carlomagno, F.; Zelano, A.; Druck, T.; Croce, C. M.; Huebner, R.
 Oncogene 9, 1591-1597, 1994
 A>Title: The human eps15 gene, encoding a tyrosine kinase substrate, is conserved in evo]


```

QY 161 GAPPVIO-----LPFAHPAATWPKSSFSRSGPSGOL-----194
D 506 GA---LQSYTGSGVYVMPQSSGPAW--PMLSFNOGLOSLTGLQPOPTGFLPSPNFS 560
QY 195 -----NTKLQKASFPVVASAPPAEAWAVQSSRLKYROLFNHDMTMSGHT 241
D 561 ATMPETAKTGFGNNELYTTSNFNNLINDSSQOKISTEKSILFYKI FETFTQNGKGLD 620
QY 242 GPQARTILMOSSLPQOALASIMWLSIDODGKLTAEFFILAMHLIDVMSGQPLPVLPP 301
D 621 SPFAVEIFRKSGLNRADLEQIMWLCDIINTGOLNKQEFALGMHLVYGLKNGKIPVLS 680
QY 302 EYIP-----PSF-----RRVSG--314
D 681 SLIPSTKLIDNLKQLTKEPTTKEKPSFGKIDALS YKNNDDVL PNYNRKRYVSAKN 740
QY 315 ---SGMSVSSSSVDQRLPEPSSDEQPEKKLPVTFEK-----KRENERSVLEKR 367
D 741 EEOSSFSPPSAKSVNH-----SSSTLQTDIDISVDKIVEKTKAPKYAGFSR---BINK 791
QY 368 RQALLQEORKEOR-----LAQLERAEOERKEREOROEAKROLE 407
D 792 NIASLEMEININPENCYDSSIPDLTSFDAI IAKPLFNEISTIDNEITNAKIOLY 851
QY 408 LEKQLEKQRELEKQREBERKEIERRBAARELOROLEWERNRROELINORKEOEGT 467
D 852 RKR---NPSSIIISGPGEITENDRKAKSRALLR-----AAMSALTGKSTESEDS 899
QY 468 VVKARKKTLFE-LEALNDKKHQLESGKODIRCLATQKQELISTNKRRLIAITHL 526
D 900 LSWEDQOQSAEIRIQOENKQOEI---IKDIRSSIDISASLSTMTGSM---ISN- 951
QY 527 QOQLOESQMLGRILPEKQILSDQLKVOQNSLHDS-----563
D 952 -QGFERWEPGIGLEDGREFL-DLKSNSNKSVTSSSPVPSSTPPVDRSSPSYSGP 1009
QY 564 LTLKALAEKELARQOLREQLDEBERETRSKLOEIDVEN--NOLKELEIHSKQOL-- 618
D 1010 KTAEBRAAYLKEQAKKMKKELAKPKNRNRNVQSSRSISSENSRQPOQIAGSSNLVBP 1069
QY 619 -----OKRLEAARLKQKQEKSELSLEKQK-----646
D 1070 RATPFQEKKEVEVAQPTQVOSTQVQPTQVQPTQVQPTQVQPTQVQPTQVQVONY 1129
QY 647 DAQRVQERDKQWLEHVOEOPRPKPHEDRLKR-----EDSVKKEAE-693
D 1130 NAKQESDDEDEDEBEKRLQELKRLKKADEKELALARKQIEDAQNESDEEFNGKD 1189
QY 694 -----RAKP-----EMQDKQSLFHPHQBPA--KLATQAPWSTTEKGPLTISA 734
D 1190 NFGHVAVVPOAAVAPASAFQSNSTNAPRSVHAAYTPAAGKNSGTGLPSTTMGNHPYFKDA 1249
QY 735 QESVKKVYVYALYRPSRSHD-EITQPGDIWVDESQCEPWLGLGELKGTGMFPANV 793
D 1250 SAS-----STSTFPAARAEMQRIQK---LDDE--DDMSDEDESSNNR---VAV 1292
QY 794 AEKIPENEVTP--AKPYTDLTSAAPKALRETAP--LPVTSSEPTTPNNWADS 847
D 1293 DNKVEEAKIGHPHARAPV---TAAPLPV---TPVPAVVPVPAANTSNKSSSIPPIA 1345
QY 848 STWSSSSNE-----KPEITMDTMAAQPSTLTVSAQLRQRS--ATPATATGSSPSVGL 901
D 1346 PIPPSVTOEPVPLAPPLPAVDGFOEPPIPSADAIATAVGKSSSPALAGVLPFP--- 1402
QY 902 QGKVEGLQALYPMRAKDNHLNFK-----SDVITVLEQDQMMWFGEVQCK 951
D 1403 --PPLPTQASTSEPIITAHVDNNGAEKGTGAYGSDSDDVLSIPES-----VGTDEEE 1455
QY 952 GWFPKSYVKLISGP 965
D 1456 GAQPVSTAGIPSP 1469

```

RESULT 9

S45781

probable calcium-binding protein YBL047c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein YBL0520

C:Species: *Saccharomyces cerevisiae*

C>Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text, change 19-Apr-2002

C:Accession: S45781; S50284; S45782; S39841; S37339; S42498

R:Goifeu, A.; Joniaux, J.L.; Purrelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, L.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45745

A:Accession: S45781

A:Molecule type: DNA

A:Residues: 1-961 <GOF>

A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c

A:Experimental source: strain S288C

R:de Wergifosse, P.; Jacques, B.; Joniaux, J.L.; Purrelle, B.; Skala, J.; Goifeu, A.

Yeast 10, 1489-1496, 1994

A:Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome II

NA-binding protein.

A:Reference number: S50284; MUID:95176707; PMID:7871888

A:Accession: S50284

A:Molecule type: DNA

A:Residues: 1-961 <DER>

A:Cross-references: EMBL:X78214

A:Experimental source: strain S288C

R:Dubois, E.; El Bakoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.;

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45782

A:Accession: S45782

A:Molecule type: DNA

A:Residues: 579-1381 <DUB>

A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c

A:Experimental source: strain S288C

R:Scherens, B.; el Bakoury, M.; Vierendeels, F.; Dubois, E.; Messenguy, F.

Yeast 9, 1355-1371, 1993

A:Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye

A:Reference number: S39824; MUID:94205266; PMID:8154187

A:Accession: S39841

A:Molecule type: DNA

A:Residues: 579-1381 <SCH>

A:Cross-references: EMBL:Z23261; NID:G313733; PIDN:CAA80797.1; PID:G313748

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:EDR1; MIPS:YBL047c

A:Cross-references: SGD:S0000143

A:Map position: 2L

C:Superfamily: yeast probable calcium-binding protein YBL047c; calmodulin repeat homology

C:Keywords: calcium binding; EF hand; transmembrane protein

F:167-199/Domain: calmodulin repeat homology <EFL>

F:560-576/Domain: transmembrane #status predicted <TMM>

Query Match

Best local Similarity 6.4%; Score 403; DB 1; Length 1381;

Matches 286; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

```

QY 3 QPPTPF-GGS-----LDVVAITVEBRKADHDOFLSLKPIAGITGDQARNFFQSGLP 54
D 109 QNPAMQSGSATGNTNNTDIPALSSNDIAKFSQLPDRTAGACQVADADKADIFLKARLP 168
QY 55 QPVAQIALADNMNDGMDQVEFSIAMKLIKLGQYQLPSTLPVWKQOQVAISSAPA 114
D 169 NQTLGEIHALCDRDASGVLDKSEFIMAMVYLQLCMSHHPMSNTPPAVLPTQ----- 219
QY 115 FGIGIASMPPLTAVAPPMKSIPIVGMSPPLVSSVPPAAVPLANAPPVIOPLPAFAH 174
D 220 -----LWDSI---RLPEPVVAVQPKRTTLPISANST 245
QY 175 PAATWPKSSFSRSGSGOLNTKLQKASFPVVASAPPAEAWAVQSSRLKYROLFNHSHDK 234
D 246 GVSLSLTHSTISRLSTGAFSNA-----ASDMSLSFEKKQOFPDAIFDLDK 290
QY 235 TMSGHLTGPQARTILMOSSLPQOALASIMWLSIDODGKLTAEFFILAMHLIDVMSGQ 294

```

```

Db      291 QHAGSLSSAVLPFPFLSRINQETLATIWDLADHNNAEFTKLEFALAMFLIOKKNAGVE 350
Qy      295 LPLPLPE-----YIPSPFRVRSGSGMSYSSSS-----324
Db      351 LPDVIPELLIOSPALGLYPNPPLPQOQSAFOIALPSRASKPSLODMPHOVSAAVNTQPT 410
Qy      325 VDQRLPE-----BPS-----SEDEQOP 341
Db      411 VPQVLPNQSNNGSLNDLLALNPFSFSPPTKAQTVQNTNTNSFYDNNNGAQLTQQOOP 470
Qy      342 EKKLPVTFEDKKRNFERSGVELEKROALLQOORKEOEL-----392
Db      471 QQPPLTHSSSGLKKTPTPS---NFGQSIIKEPBEQOLRESSDTSFSAQPPVPKXAS 526
Qy      383 -----382
Db      527 SPVKTASTTLTPQVPNFSVSPMPAGATSAATGAAGAAGAAAGASAFSSSNNAFKN 586
Qy      383 -----AQLERAEOEREREROEAKRQOLEKOLEKORELEROREEERKEIER 432
Db      587 QDLFADGEASQSLSNATTETMANLSQVNSLSKQASITNDKKSRAQOLKKTETKXSIQI 646
Qy      433 REAKRELERORELERERNRQOELLNQKOEQTVLKKARKTLEPELALNDKKHOLE 492
Db      647 KLNLRSTHDQNVKOTDLEAQLV--QVKNEN-----TLAQOLVSEANTHAAE 694
Qy      493 GKLDICRRLATOROIESTNKSRELIAEI---THLOOQLOSOQMLGRLLPEKQILS 548
Db      695 SKINE---LTTDQESQTKVAELKEQITNLMSWTASLOSQLEKQO-----737
Qy      549 DOLKQOVQNSLHRDSLTLTKRALFAKELARQOLEQDEVERETRSKLOEIDVPNNOLKE 608
Db      738 -QVQKQ---ERSMVDNSKQLELNOVTYANLQKEIDGG-----EKISVYLTKQKE 793
Qy      609 LREHSKQOOLOKORSLERARKQOEOE--RKSLLE---EKQKEQACRRVOERDKOMLEHV 663
Db      784 LND-----YKTYEBQHAQLOAKYQDLSNKDITDLTREKQLEBRNRNQIEQENLHYQH 837
Qy      664 QOEOPRRKREHEDRLKREDSVRKKEAEERAK-----PENQDYQSKLPHPHQEP-- 713
Db      838 SKLOE---MPDLSQKASFEKADQELKERNIEYANNVRELSEKQML--AMGQLPED 890
Qy      714 -----AKLATAQPMST---TEKGPL---TIS---AQESKVVYVYALVPFERSHDEI 757
Db      891 AKDIIASASNTDTTTEKATSRGVNHEHTYKFEVETTENSNNLVNRVKODEETERTES 950
Qy      758 TIQPGDI--VWVESQTEGPPWLGGELKQKGTGMPANYAEKIPE-----NEVPTP--AKPV 809
Db      951 DVFPRDVTLLSQSDSEBANTNNGTQSGNETANP--NLTEITLSDRFDGDLNEYGIPRSQSL 1009
Qy      810 TDLISAPKALARETPAPLEVTSEPEPTTNMADSSSTPSSSNKEPFDNDMDTAAQ 869
Db      1010 TSSVANNAPQ--SVRD--DVELPETLEERDTI--NNTANRNT---GNLSHPGEAEATPAT 1062
Qy      870 PSLTPVSAQURQSAFTPATATGSSPEPVLG--QGEKEVEJQAQ--ALYPMWAKKDN- 923
Db      1063 ASTDVL-----NETTEVLEDSGTTGRANSNEDGEVSSIQSPKISAP--KAKTINE 1114
Qy      924 -----HUNFKSDVITVLEQODMMWGEVOGQKGMFPKSYVKLISGP-----965
Db      1115 EFPPIQELHIDESDSSSDDE---FEDTRE---IPSAIVKTLQTPYNAQPTSSLEIH 1166
Qy      966 ---VRKSTSIDQTPESASIKRYVASPAKAPAI--GEPIFAMTYTSSSEGDLTFQGD 1020
Db      1167 TEQVIKTPAPGTSBHNENSKASTNSILPVKQEFDDFAGL-----EQAAVEEDNG- 1219
Qy      1021 VIIVTKKDDMMWTGTVGSKGVFPSPNYVRLKDS-----GSGTAGKTSIGKKPEI 1071
Db      1220 -----ADSESEFEFVANAGSMEGETIDHKDLDELQMAFTGLTSSSNPTI--PKQV 1272
Qy      1072 AQVIASVATGPBQULTLAPGQILIRKKNPGGWEGEIQAGKKRQICGMFPANYVKLSP 1131

```

```

Db      1273 QQ-----QSTSDPAQV-----SNDEND-EIFA-----GF-----1295
Qy      1132 GTSKITPTTEL---PKTAVQPAVCQVIGWYDTAOND--ELAFSKQGIIN 1176
Db      1296 GNSKAEPTKVAFTPSIQOP-----IPLKNDPIVDASLSKGPVIV 1334

RESULT 10
S28589
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: S28589
R:Piecz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <PIE>
A:Cross-references: EMBL:Z19092; NID:91746; PIDN:CAA79519.1; PID:91747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she
Covalent modifications to this protein include conversion of arginine to citrulline and t
C:Genetics:
A:Introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match      6.2%; Score 391.5; DB 1; Length 1407;
Best Local Similarity 29.7%; Pred. No. 1.2e-08;
Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

Qy      326 DQRLPEPSSSEBOQPEKKLPVTFEDKKRNFERSGVELEKROALLQO-----375
Db      198 EEFIEEQQLRRRQQLKRELREEQQRERREOHERALOEEBQQLRQRWREBPBQ 257
Qy      376 -----RKEORLAOLERAEOERKEREROEAKRQOLELEKOLEKORELEROREEER 426
Db      258 QLRLELEIREREOQLRBOERREOQLRBOERREOQLRBOERREOQLRBOERREOQLRBOER 317
Qy      427 RRE-----IERRE-AAKREL---ERQOLEMERNRROELNQRNKEO--EGVIVLKAR 473
Db      318 RREOQLBOERREOQLKRELEIRERREOERLEO--ERRQLLAEVRERARERGSGLRRW 376
Qy      474 RKTLEPELALNDKKHOLEGKLDICRRLATOROIESTNKSRELIAEITHLQOOLQ-- 531
Db      377 QRLSEAGARQSKVYSPRRROEBSLPODERKO---RQERRELEQARRQOQMAE 432
Qy      532 -ESQMLGRL-----IPEKQILSDQKQVOQ---NSLRDSTLTKRALFAKELARQ 580
Db      433 EESERRQQLSARPSLRERQRLAEREOEOEPFREERQERREOQLQFLREBOLQRE 492
Qy      581 LREQLDE-----VERETSKQOET---DVFNQOLKE--LREIHS-----KOOLQKRS 623
Db      493 RAQQLQOEEDSFQEDERRRROEOQRPQGTWQLOEABORRHHTLYAKGQOEBQREEE 552
Qy      624 LEARLKQOEKRSKLELEK--QKEDARVVOERDKQW--LEHVQOEOPRRKPHEDR 679
Db      553 LOBEKRRQ--ERRREYREBEKLOREDEKRRQERROVRLBELRLQEBQDLDRKLREBQ 611
Qy      680 L--KREDSVRKKEAEERAKPEMQ 700
Db      612 LIQEREERERLRQERERERKLREBQ 635

RESULT 11
A45592
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jun-2000
C:Accession: S24597; A45592; S24592; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.

```


submitted to the EMBL Data Library, November 1990

A:Reference number: S24597

A:Accession: S24597

A:Molecule type: DNA

A:Residues: 1-1909 <ZHU>

A:Cross-references: EMBL:X56203; NID:g9915; PID:g9916

R:Zhu, J.; Hollingdale, M.R.

Mol. Biochem. Parasitol. 48, 223-226, 1991

A:Title: Structure of Plasmodium falciparum liver stage antigen-1.

A:Reference number: A45592; MUID:92107224; PMID:1840628

A:Accession: A45592

A:Molecule type: DNA

A:Residues: 1-195;638-688;1165-1215;1590-1909 <ZHU>

A:Note: Sequence extracted from NCBI backbone (NCBI:75012, NCBI:75014, NC

R:Guertin-Marchand, C.; Drullhne, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudin,

Nature 329, 164-167, 1987

A:Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c

A:Reference number: S29393; MUID:87315391; PMID:3306406

A:Accession: S29393

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 323-387 <GUE1>

A:Cross-references: EMBL:M28266

R:Guertin-Marchand, C.; Drullhne, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudin,

submitted to the EMBL Data Library, April 1992

A:Description: a liver-stage-sepific antigen of plasmodium falciparum characterized by

A:Reference number: S34842

A:Accession: S34842

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 323-381, 'HKAT' <GUE2>

A:Cross-references: EMBL:M28266

A:Note: difference at carboxyl end due to frameshift error

C:Comment: This protein is found as flocculent material in the parasitophorous vacuole.

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: EF hand

F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Query Match

Best Local Similarity 28.3%; Score 383.5; DB 2; Length 1909;

Matches 164; Conservative 113; Mismatches 198; Indels 105; Gaps 27;

Qy 197 KLQAKQSPVNASAPPAEAVNPQSSRLKYRLFNHDKTMSGHLTGPPARTIMOSSLPQ 256

Db 1041 KLQEQQS-DLEQRLAKKLEQEQSDLEQRLAKE-----KLQ-----QSDLEQ 1085

Qy 257 AQLAS---IWNLSIDIDQ---GKLTAEFTLAM-HLIVANSQPLPVLPPEYIPPSF 308

Db 1086 ERLAKKLEQEQSDLEQRLAKKLEQEQSDLEQRLAKKLEQ-----QSDL 1134

Qy 309 RRVSSGMSVSISSSSVDQ---RLPEPSS-EDEQPEKKLPVTFEDKKRENFERS 361

Db 1135 EQLRLAKKLEQEQSDLEQRLAKKLEQEQSDLEQRLAKKLEQEQSDLER--TKAS 1191

Qy 362 VLEKRRQALLLEQORKEOERL-AQLERAEERKERER-EOEA---KRLLELEKLEKOR 416

Db 1192 KETLQEQSDLEQRLAKKLEQEQSDLEQRLAKKLEQEQSDLEQRLAKKLEQEQS 1251

Qy 417 ELERQRE---EERKELEERRAEAKREL-ERQRLEREMERNRQELINPNK-EQGTIV 469

Db 1252 DLEQERAKKLEQEQSDLEQRLAKKLEQEQSDLEQERAKKLEQEQSDLEQERLAK 1311

Qy 470 LKARRTLLEFELALNDKQKQLEKQDRIKRLATQROEISTNKSRELRLAETITLQOQ 529

Db 1312 EKLQEQSDLEQRLAKK-----LQEQSDLEQERRAKKLEQEQSDLEQ 1358

Qy 530 -----LQEQSDLEQERRAKKLEQEQSDLEQERRAKKLEQEQSDLEQ 583

Db 1359 RLAKKLEQEQSDLEQERRAKKLEQEQSDLEQERRAKKLEQEQSDLEQERRAKKLEQ 1418

Qy 584 QLDEVERETRS--KLQETIVFNQKEL-REHSKQOLQKRS--LEAPRLKQEKDERKS 638

Db 1419 QQSDLEQERRAKKLEQ-----QSDLEQERRAKKLEQEQSDLEQERRAKKLEQEQS 1472

Qy 639 -LE---LEKQEDAKRRVQRRQKQWLEHVQOE---EQPRPKPH-----EEDRL-- 680

Db 1473 DLEQRLAKKLEQEQSDLEQERRAKKLEQEQSDLEQERRAKKLEQEQSDLEQERRAN 1532

Qy 681 -KREDSVKKAEERAKPEMODKOSRLFNHQBPAKLATQ 719

Db 1533 EKLQEQSDLEQERRAKKLEQEQSDLEQERRAKKLEQEQ 1572

RESULT 12

A45973

Trichohyalin - human

C/Species: Homo sapiens (man)

C/Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C/Accession: A45973

R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.

J. Biol. Chem. 268, 12164-12176, 1993

A:Title: The structure of human trichohyalin. Potential multiple roles as a functional E

ed (cross-linking) protein.

A:Reference number: A45973; MUID:93280194; PMID:7685034

A:Accession: A45973

A:Molecule type: DNA

A:Residues: 1-1898 <LEE>

A:Cross-references: GB:U09190; NID:g292835; PIDN:AA65582.1; PID:g292836

A:Date: authors translated the codon AGC for residue 1714 as Pro

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she

valent modifications to this protein include conversion of arginine to citrulline and t

C:Genetics: GDB:THH

A:Gene: GDB:THH

A:Cross-references: GDB:136223; OMIM:190370

A:Map position: 1921-1921

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF>

Query Match

Best Local Similarity 27.0%; Score 381; DB 1; Length 1898;

Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

Qy 326 DQRLPESSSDLEQPEKKLPVTFEDKKRENFERSYELKRRQALLLEQ--RKEQ--R 381

Db 348 EQQLRE---QEERRRQQLRREQEERRRQQLRREQEERRRQQLRREQQLR 404

Qy 382 LAQLERAEERKERE-ROEQEAKR--OLELEKOLEKRELEPQREER----- 426

Db 405 EQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQEERRHOKHQBERRQ 464

Qy 427 --RKEIERREBAKRELEOROLEWERNRQELNQRNKEQGTIVLKARKTLLEFLEA 483

Db 465 RLKREQEERRRQWLKREETERH---EQERRQQL-KDQEE-----RERWLKLEE- 513

Qy 484 LNDKQKLEKQDRIKRLATQROEISTNKSRLRLAETITLQOQEQO-----QML- 537

Db 514 --ERREQERREQQLR-REQEERRRQQLRREQEERRRQQLRREQQLRREQEERLEQLK 570

Qy 538 -----GRLPEKQILSDQ---KQVQNSLHRSLSLTAKALAKELARQ 580

Db 571 REEKRLQERREQQLRREQEERRRQQLRREQEERRRQQLRREQEERRRQQLRREQEERLE 630

Qy 581 LREQLDE-VERETRSKLEQIDVFNQKELR--EHSKQOLQKRSL-----EAPRLKQK 632

Db 631 QEERDEHLKKEEPEERHRLHLSQEGERHRLHLSQEGERHRLHLSQEGERHRLHLSQE 690

Qy 633 -----EQERKLEL-EKQEDAKRRVQRRQKQWLEHVQOEQPRPK----- 673

Db 691 LKREHEERRRQQLAEQEQERAKRIRKSRIPKQWQLESADAAQSVLLLEAPQAGRAEA 750

Qy 674 PHEEDRLKREDSVKKAEERAKPEMODKOSRLFNHQBPAKLATQAPWSTTEKGPLTIS 733

Db 751 PQEQEKRRRSSELOQEEERAHNQQQEEBRKPLTWQMAE-----EKSEGRQRSL 803

Qy 734 AQESKVVVYRALYVPFSRSHDE 756

Db 804 ARPPUREQRERQOLRAEEQRQREQ 826

RESULT 13

A40691
trichohyalin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C/Accession: A40691, A34209, S32633
R/Fietz, M.J.; McLaughlan, C.U.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A>Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding
A/Reference number: A40691, MUID:93260018, PMID:7684041
A/Accession: A40691
A:Molecule type: DNA
A/Residues: 1-1549 <FIE>
A/Cross-references: EMBL:Z18361, NID:g295940, PIDN:CAA79165.1, PID:g295941
A/Note: sequence extracted from NCBI backbone (NCBI:P132511)
R/Fietz, M.J.; Prestland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A>Title: The cDNA-decoded amino acid sequence for trichohyalin, a differentiation marker
A/Reference number: A34209, MUID:90130632, PMID:2298812
A/Accession: A34209
A:Molecule type: mRNA
A/Residues: 1016-1151, 1205-1257, 1281-1398, G', 1400-1549 <F12>
A/Cross-references: GB:X51695, PIDN:CAA5992.1, PID:g1828
A/Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covariant modifications to this protein include conversion of arginine to citrulline and
C/Genetics:
A/Intons: 46/3
A/Note: single copy gene
A/Superfamily: trichohyalin, calmodulin repeat homology
C/Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F/49-81/Domain: calmodulin repeat homology <EF2>
F/387-851/Region: 28-residue repeats
F/886-1519/Region: 23-residue repeats

Query Match 6.0%; Score 375.5; DB 1; Length 1549;
Best Local Similarity 27.6%; Pred. No. 5.9e-08;
Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;

332 EPSEDEQPPKTLPTFE-----DKRPNFSGVLEKRRQALDEQKROEQLAQLR 387
Db 383 KPQREBYREBEQRLKAEKQLQREKRQREKQYREVLQRE--ERQREBEQLQREER 440
QY 388 AEQERKERQ-----DOFAKQOLEKQ-----LEKQRELERQREER 426
Db 441 EKRRQREKQYLEKVELMEBEQQLQREBERKQREKQYLEKVELMEBEQLQREERK 500
QY 427 RKELERRAAKRELERQQLMEWERNRQELLNQRNKEQGVVLKARKTLFELEALND 486
Db 501 ROEBERQYLEVELQEEQLQREERK-----RRQRE-----ROYLE----- 538
QY 487 KKHOLEGLQDIRLQATQROEISTNKSRLIAETIHLQ--QOLOESQOMLRLIPEK 544
Db 539 -KVELQEEQ-----LQREBERKQREKQYLEKVELQEEBQLQREKQREK 590
QY 545 QILSDQLKQVQNSLHRDSSLTLKRALFAKELAQQLREQLDEVERTRSKLQEIIVFN 604
Db 591 QYL-EKVELQEEBQLQREBERKQREKQYREVLQEE--EQVQROBERK-----RR 641
QY 605 QLKELREIHSQQLQKQSLAARLKQKQERKSLJLEK-----OKEDAQR 650
Db 642 QERERQYLEKQLEQEEBQLQREBERKQREKQYREVLQEEBQLQREERK 701
QY 651 RVQERDKQWL--EHVQEQPRPKPHEDRLKEDSVRKKAER--AKPMQDKSRL 706
Db 702 ROEBERQYLEKELQRE--RLQREKQLEQREDEKRRQVVERKYVEELQEEBRL 758

RESULT 14
T15597

hypothetical protein C25A11.4b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T15597
R/Favella, T.
submitted to the EMBL Data Library, October 1995
A>Description: The sequence of C. elegans cosmid C25A11.
A/Reference number: Z18375
A/Accession: T15597
A>Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A/Residues: 1-737 <FAV>
A/Cross-references: EMBL:U39650, NID:g1049376; PID:g1049381; PIDN:AAA80392.1; CESP:C25A11
A/Experimental source: strain Bristol N2
C/Genetics:
A/Gene: CESP:C25A11.4b
A/Intons: 65/3; 86/3; 126/3; 433/2; 494/2; 711/2

Query Match 5.7%; Score 359.5; DB 2; Length 737;
Best Local Similarity 23.6%; Pred. No. 1.1e-07;
Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;

QY 181 KSSFSRSGPQGLNTKQKQSPVYASAPPAEYVQSSRLKYROLFNS-----HDKT 235
Db 7 QSISSREDLSEHATSSVREIPVHRAPSTAP--SHSSVFEKHMPPTTSTYHHVET 63
QY 236 MSG--HLNPGQARTIMOS--SLPQALASIMNLSIDIDQDKLTAEFFILAMHLIDVAMS 291
Db 64 PDEXYRRNRVMTITITSTALSGTLP----- 91
QY 292 GQPLPVLPPEYIP-PSRRVRSGSGMSVISSSYDQRL-----PEEESDE--Q 339
Db 92 GRPASPL--DRLLPYPTITTTTGGGRT-REKTVDYVYTHRIDIEQERRIRDOARQ 148
QY 340 QPEKLPVTFEDKKEENRSGVLEKRRQ-----ALLQQRKEQRLAQLEAEQER 392
Db 149 QEEQRRREDNARILILAQREHQMERLRQEOONLSERLALREKADKRLQ--RLLRQ 207
QY 393 KEREQEQ-----EAKQOLEKQLEKQRELERQREERKEIERREAAKRELER 442
Db 208 REKKREEDWDLSEIIRLAEAEALRRALREKER-IDREKAEERKMTERRERARLER 266
QY 443 QQLMEWRNRQ-----ELLNQRNKEQGV--VLKARKTLFELEALNDKQKQLEK 495
Db 267 ER-LEBERQKAKATTEIERRRHERIEIERIRKREIERRRERREKKAED-- 322
QY 496 QDIRQLATQROEISTNKS--RELRIAEITHLOOQLOESQOMLRLIPEKQILSDQLKQV 554
Db 323 ---RLRLERLEIRIERRELERERQRELELQREAEER-----ERQLEDEAREM 371
QY 555 -QONSILHRDSSLTLKRALFAKELAQQLREQLDEVERTRSKLQEIIVFNQ--KELREI 612
Db 372 RRREBRRAELVADVHRQAEERERLRQREERERLERIRLQDKLDMERIDARER 431
QY 613 HSKQOLQKQSL--FAARLKQKQERKSL--ELEKQEDAQRVQERDKQWLHVQEEQ 668
Db 432 EKKEERRREFELIEAARKKEARDRLDEMEREVRREERREKERRQERRIAEKER 491
QY 669 PRPKRPHQ-----DLKREDSVKKAEERAKPEMODKQSL 706
Db 492 KRQREBEETIARINELQRAAARQQRNAELDRQORDELRK--AQELSEREMKEKERD 550
QY 707 PPHQEPALQATQAWSTTEKQPLTISAQESKVYVYALPYFSSRSHDETITQGDIVM 766
Db 551 RERANDEAQLALLERHNQILIRENERREAVANNRLE--DRRSQDL-----DHIV 603
QY 767 VDESQTGEFGWLGGELKQKGTGFPANVAEKIPENEVPTPAKVTDLTSAPAKLARETP 826
Db 604 RERSEKQ-----FELEKRRLLAEKAMARKK--NHLTSS-----TLAKLT 643
QY 827 APLPVTSSEPTTP-----NNWADPSSTWSSS 854
Db 644 QPMYYTTRREPVTKVERQVIERIDRNWVEDVYAPQS 683

RESULT 15
T15598
hypothetical protein C25A11.4a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 20-Sep-1999
C/Accession: T15598
R/Favella, T.
submitted to the EMBL Data Library, October 1995
A/Description: The sequence of C. elegans cosmid C25A11.
A/Reference number: Z18375
A/Accession: T15598
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1017 <FAV>
A/Cross-references: EMBL:U09650; NID:g1049376; PID:g1049380; PIDN:AAA80391.1; CESP:C25A1
A/Experimental source: strain Bristol N2
A/Genetics:
A/Gene: CESP:C25A11.4a
A/Intons: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2; 735/1; 792/3; 833/3; 873/2

Query Match 5.7%; Score 359.5; DB 2; length 1017;
Best Local Similarity 23.6%; Pred. No. 1.6e-07;
Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;

QY 181 KSSSFSSGSGSOLTKLQKQSFVVASAPPAEWAVPOSSRLKYROLFNS-----HDKT 235
DB 7 QSISSREDLLSEHATSRSTVREIPVHRAPSTAP---SHSVFEYHMPPTTSTYHHVET 63

QY 236 MSG--HITGQARTLMQOS--SLPQQLASIMNSIDIDQDKLTAEFFILAMHLIDVANS 291
DB 64 PSDEYRREVMTRITITRSTALSTPL----- 91

QY 292 GQPLPVLPEEYIP--PFRFRVRSQSGMSVSSSDQRL-----PEEPPSEDE---Q 339
DB 92 GRPASPL--DRYLPYPTTTTSSGDGT--REKTVDTKYVTHRDIEQERRIREDOARRQ 148

QY 340 QPEKKLPVTEFDKKRENFERSVLEKRRQ-----ALLEQQRKEOERLAQLERAEOER 392
DB 149 QEEQDRDRREDNARRILAQREHQMRLRQQLNSERALARERADKERLQ--ERLLRQ 207

QY 393 KERERQO-----EAKRQLELEKOLEKORELEROREBERKEIERREPAKRELER 442
DB 208 REKKRREEMRLSIRLVAEEAEELARRALEKER-IDREKAEERKTMERLERERARLER 266

QY 443 QROLEMERNRQ-----ELNQNKQEGTV--VLKARKTLFELEALNDKKGQLEGKL 495
DB 267 ER-LEERRQKEKAEYERIEREREREHERIERIERIKRERIREREREKKAED--- 322

QY 496 QDIRCLATOROETESTNKS--RELRIAEITHLOOQLQESQOQLGRLLPEKQLSDQLQV 554
DB 323 -----RLRRLERLEERIERERRELEAREQELQRRREADR-----EQRLDEAREM 371

QY 555 -QONSLSHDSLLTKRLAEKELARQQLREQLDVERETRSLQEIQIDVENNQ--KELREI 612
DB 372 RRREERERREAEADVRAQAEERERLAKQREBEAERLERIRLEQOKIDMERIDAEERER 431

QY 613 HSKQOLQKQSL--EAARLKQKQEKSL---ELEKQKEDAQRRVQRDQKMLEHVQEEQ 668
DB 432 ERKEBERREFELIEAARKKEARDRLDEMERERVAEEEREREKERRERQERRIAEKER 491

QY 669 PRPRKPRHE-----DLKREDSVKKKEAEERAKPEMODKOSRL 706
DB 492 KRQOESEEFELARINELORAAARQAQONALDQROQDELDRK-AQELSEREMEKERKD 550

QY 707 FHPHQEPAKLATQAPWSTTEKGLTISAQESKVYVYRALYPRESSGSHDEITIQPGDIYM 766
DB 551 RERANBEAQLADLERERHNLIRENERREAVERRANNRLE--DRSRDYL-----DHIV 603

QY 767 VDESQTEGPEMGLGELKGTGWPPANVAEKIPENEVPTPAKPYTDLTSAPAKLALRETP 826
DB 604 RERSEKEQ-----FELEKRLLAEKEAMNRKK--NHLISSE---TLAKLT 643

QY 827 ADLPVTSSEPTTP-----NNWADFSSSTWSSS 854
DB 644 QPMYTTREPEVTTKVERQVIERIDRNWVVEDVPYAPSQS 683

Search completed: March 14, 2003, 12:08:52
Job time : 36 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 08:18:17 ; Search time 38 Seconds

(without alignments)
1323.968 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269
Sequence: 1 MAQPTPGSLDVMATVE.....VGLFPSNYVKLTDMPSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6269	100.0	1714	1 ITN1_MOUSE	Q920R4 mus musculu
2	6021	96.0	1217	1 ITN1_RAT	Q9wwe9 rattus norv
3	5842.5	93.2	1721	1 ITN1_HUMAN	Q15811 homo sapien
4	5063.5	80.8	1270	1 ITN1_XENLA	Q42287 xenopus lae
5	3068	48.9	1696	1 ITN2_HUMAN	Q9nzm3 homo sapien
6	3049.5	48.6	1658	1 ITN2_MOUSE	Q920R6 mus musculu
7	486.5	7.8	896	1 EP15_HUMAN	P42866 homo sapien
8	466	7.8	897	1 EP15_MOUSE	P42867 mus musculu
9	410.5	6.5	1480	1 PAN1_YEAST	P32521 saccharomyc
10	403	6.2	1381	1 YBET_YEAST	P34216 saccharomyc
11	391.5	6.2	1407	1 TRHY_RABIT	P37709 oryctolagus
12	381	6.1	1988	1 TRHY_HUMAN	Q07283 homo sapien
13	375.5	6.0	1549	1 TRHY_SHEEP	P22793 ovus aries
14	330	5.3	793	1 CALD_HUMAN	Q05682 mus sapien
15	327	5.2	1794	1 YAVI_SCHPO	Q10172 schizosach
16	314.5	5.0	887	1 YLX8_CAEEL	P46504 caenorhabd
17	306.5	4.9	771	1 CALD_CHICK	P12957 gallus gall
18	304.5	4.9	585	1 INVO_HUMAN	P07476 homo sapien
19	304.5	4.9	958	1 IF3A_TOBAC	Q40554 nicotiana t
20	302	4.8	4684	1 PLEI_HUMAN	Q15149 homo sapien
21	298	4.8	4473	1 PLEI_CRIGR	Q91555 cricetulus
22	296	4.7	4687	1 PLEI_RAT	P30427 rattus norv
23	283	4.5	733	1 VINE_MOUSE	Q91128 mus musculu
24	283	4.5	1805	1 NEST_RAT	P21263 rattus norv
25	281.5	4.5	544	1 INVO_AOTTR	P24708 aotus trivi
26	280	4.5	522	1 INVO_HYLLA	P17941 hylobates i
27	280	4.5	1938	1 MYS_AEQIR	P24733 aequipeten
28	279	4.4	678	1 GLT_DROME	P33438 drosophila
29	278.5	4.4	377	1 GARP_PLAUF	P1816 plasmodium
30	277	4.4	1300	1 NCK1_HUMAN	P16333 homo sapien
31	276.5	4.4	407	1 DYNA_NEUCR	Q01397 neuropa
32	276	4.4	407	1 M21_STRPY	P50468 streptococc
33	274.5	4.4	2230	1 GOG4_HUMAN	Q13439 homo sapien

34	273.5	4.4	671	1 VINE_HUMAN	Q60504 homo sapien
35	273.5	4.4	857	1 NFW_CHICK	P16053 gallus gall
36	273.5	4.4	1344	1 IF3A_MOUSE	P23116 mus musculu
37	273.5	4.4	1382	1 IF3A_HUMAN	Q14152 homo sapien
38	273	4.4	1976	1 MYHA_BOVIN	Q27991 bos taurus
39	273	4.4	1976	1 MYHA_RAT	Q91100 rattus norv
40	272.5	4.3	501	1 MYSU_RABIT	Q99105 oryctolagus
41	272	4.3	560	1 INVO_PANPA	P14591 pan paniscu
42	272	4.3	1123	1 DC11_DROME	P18169 drosophila
43	271.5	4.3	2245	1 MYSJ_DICDI	P54697 dicystostel
44	271	4.3	835	1 INVO_PONPY	P14708 pongo pygma
45	271	4.3	1976	1 MYHA_HUMAN	P35580 homo sapien

ALIGNMENTS

RESULT 1
ID ITN1_MOUSE STANDARD; PRT; 1714 AA.
AC Q920R4; Q9R143; 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Intersectin 1 (EH and SH3 domains protein 1).
GN ITSN1 OR ITSN OR ESE1.
OS Mus musculus (Mouse).
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP MEDLINE=99164083; PubMed=10064583;
RX Sengar A.S., Wang W., Bishop J., Cohen S., Egan S.E.;
RT "The EH and SH3 domain ESE proteins regulate endocytosis by linking to
RT dynamin and Eps15.";
RL EMBO J. 18:1159-1171 (1999).
[2]
RN SEQUENCE OF 966-1714 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 545-599
RP FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/Ola; TISSUE=Splice;
RA Skripkina I.Y., Tsyba L.O., Anoprienko O.V., Slavov D., Tassone F.,
RA Ryndlich A.V., Gardiner K.;
RT "Mouse homologues of human chromosome 21 genes.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Adapter protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -!- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters
CC several dynamin in a manner that is regulated by alternative
CC splicing. Also binds clathrin-associated proteins and other
CC components of the endocytic machinery, such as N-WASP, Eps15 and
CC Stc12 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC Enriched in synaposomes (By similarity).
CC -!- ALTERNATIVE PRODUCTS: At least 2 isoforms, 1/Ese1 (shown here)
CC and 2; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Widely expressed. Expressed at high levels in
CC brain, heart and skeletal muscle.
CC -!- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,
CC bind to dynamin (By similarity).
CC -!- DOMAIN: The KLERO domain binds to SNAP-25 and SNAP-23 (By
CC similarity).
CC -!- MISCELLANEOUS: Overexpression results in the inhibition of the
CC transferrin uptake and the blockage of the clathrin-mediated
CC endocytosis.
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EH DOMAINS.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF132481; AAD19749.1; -;
 DR EMBL; AF132478; AAD19746.1; -;
 DR EMBL; AF169621; AAD48848.1; -;
 DR EMBL; AF356517; AAK40228.1; -;
 DR HSP; P29354; IGPC.
 DR MGD; MG1.1338069; Itsn.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_repeat.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; Rhogef.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 5.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; Rhogef; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 5.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00054; Efh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; Rhogef; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS50004; C2_DOMAIN_2; 1.
 DR PROSITE; PS50010; DH_2; 1.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 5.
 DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 221 310 EH 2.
 FT CA BIND 267 279 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
 FT DOMAIN 352 662 COILED COIL (POTENTIAL).
 FT DOMAIN 738 799 SH3 1.
 FT DOMAIN 906 964 SH3 2.
 FT DOMAIN 995 1053 SH3 3.
 FT DOMAIN 1067 1131 SH3 4.
 FT DOMAIN 1148 1207 SH3 5.
 FT DOMAIN 1230 1416 DH.
 FT DOMAIN 1455 1564 PH.
 FT DOMAIN 1576 1672 C2 DOMAIN.
 FT DOMAIN 321 324 POLY-SER.
 FT VARSPIC 1214 1714 MISSING (IN ISOFORM 2).
 SEQUENCE 1714 AA; 194284 MW; 4D7AF298397860A7 CRC64;

Query Match 100.0%; Score 6269; DB 1; Length 1714;
 Best Local Similarity 100.0%; Pred. No. 1e-237;
 Matches 1213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAOPEPTPGGSLDWAIIVERAKHDOOFLSLKPIAGITGSDOARNFPPQSGLPQVLAQ 60
 Db 1 MAOPEPTPGGSLDWAIIVERAKHDOOFLSLKPIAGITGSDOARNFPPQSGLPQVLAQ 60

QY 61 IVALADNNDRMDQVEFSIAMKLIKLGQYQULPSTLPVPMKQOPIVAISSAPAFGIGI 120
 Db 61 IVALADNNDRMDQVEFSIAMKLIKLGQYQULPSTLPVPMKQOPIVAISSAPAFGIGI 120

QY 121 ASMPPLTVAAPVPMGSLFVVGMSPLVSVPPAAVPLANGAPVIOPLAFAPATWP 180
 Db 121 ASMPPLTVAAPVPMGSLFVVGMSPLVSVPPAAVPLANGAPVIOPLAFAPATWP 180

Db 121 ASMPPLTVAAPVPMGSLFVVGMSPLVSVPPAAVPLANGAPVIOPLAFAPATWP 180

QY 181 KSSFSRSGQSGSLWTQKQAKSPVAGAPPAEAAVQSSRLKYROLFNHDKTMGHL 240
 Db 181 KSSFSRSGQSGSLWTQKQAKSPVAGAPPAEAAVQSSRLKYROLFNHDKTMGHL 240

QY 241 TGPQARTIMOGSLPQAOALASIMLSIDIDQCKLTAEFFILAMHLIDVANGQPLPVLP 300
 Db 241 TGPQARTIMOGSLPQAOALASIMLSIDIDQCKLTAEFFILAMHLIDVANGQPLPVLP 300

QY 301 PEYIPSPFRVAVSGSGMSVSSSVQDRLPEPSSDEQPEKKLPVTFPEKKENFERG 360
 Db 301 PEYIPSPFRVAVSGSGMSVSSSVQDRLPEPSSDEQPEKKLPVTFPEKKENFERG 360

QY 361 SVELEKRRQALLQQRKEERLAQERAEQERKERERQEQEAKQLELEKQERELER 420
 Db 361 SVELEKRRQALLQQRKEERLAQERAEQERKERERQEQEAKQLELEKQERELER 420

QY 421 QREERERKEIERREAKRELERQERLEWERNRQELNQRKKEGTVVLKARKTLEFE 480
 Db 421 QREERERKEIERREAKRELERQERLEWERNRQELNQRKKEGTVVLKARKTLEFE 480

QY 481 LEALNDKKGHEGKQDRIKRLATQREIESTNKSRELIAETHTLOOLQESQOMGL 540
 Db 481 LEALNDKKGHEGKQDRIKRLATQREIESTNKSRELIAETHTLOOLQESQOMGL 540

QY 541 IPEKOILSDOLKQVOQNSLHRDSLITLKRALFAKELARQUREOLDREVERETRSKLOEID 600
 Db 541 IPEKOILSDOLKQVOQNSLHRDSLITLKRALFAKELARQUREOLDREVERETRSKLOEID 600

QY 601 VFNNOLKELREIHSKQOLQOKRSLAARLKQKEBRKSLLEKQEKDQORVQSRDKWL 660
 Db 601 VFNNOLKELREIHSKQOLQOKRSLAARLKQKEBRKSLLEKQEKDQORVQSRDKWL 660

QY 661 EHVOQEOGRPRKPEHEDRLEKREDSVRKKEAEERAKPEMODQKSLFPHQEPKALATQA 720
 Db 661 EHVOQEOGRPRKPEHEDRLEKREDSVRKKEAEERAKPEMODQKSLFPHQEPKALATQA 720

QY 721 PWSTTEKGLTISAQESKVVYRRLYPFESRSHDEITIQGDIIVMDESGTGPBGWG 780
 Db 721 PWSTTEKGLTISAQESKVVYRRLYPFESRSHDEITIQGDIIVMDESGTGPBGWG 780

QY 781 ELKGTGMPANVAKIEIPENEVPTPAKPVTDLTSAAPKALARETPALPTTSSEPTTP 840
 Db 781 ELKGTGMPANVAKIEIPENEVPTPAKPVTDLTSAAPKALARETPALPTTSSEPTTP 840

QY 841 NNWADPSSSTWPSSENEKEPTDNDTMAAQPULYPSAGQLFORSAFTPATITGSSPSV 900
 Db 841 NNWADPSSSTWPSSENEKEPTDNDTMAAQPULYPSAGQLFORSAFTPATITGSSPSV 900

QY 901 GQGEVVEGLQAOALYPMRAXKDNHLENFKSDVITVLEQDDMMWGEVQOGKGFPPKSVK 960
 Db 901 GQGEVVEGLQAOALYPMRAXKDNHLENFKSDVITVLEQDDMMWGEVQOGKGFPPKSVK 960

QY 961 LISGPRKSTSIDTPTSPASLKRVASPAKPAIPGEEFIAMYTESSBOGDLTFQGD 1020
 Db 961 LISGPRKSTSIDTPTSPASLKRVASPAKPAIPGEEFIAMYTESSBOGDLTFQGD 1020

QY 1021 VIVVTKKQDMMWTGTVGKSGVFPNSNYRLKDSBSGAGTGSIGKKPEIAQVIASTAA 1080
 Db 1021 VIVVTKKQDMMWTGTVGKSGVFPNSNYRLKDSBSGAGTGSIGKKPEIAQVIASTAA 1080

QY 1081 TGPQOULTAPGQLIRKKNGCGMWEGELQARGKKRQIGWFPANVVKLISGTSKITPTE 1140
 Db 1081 TGPQOULTAPGQLIRKKNGCGMWEGELQARGKKRQIGWFPANVVKLISGTSKITPTE 1140

QY 1141 LPTKAVQPAVCVIGMYDTAQNDELAFSGQIINVLNKDPPMWKGEVSGQVGLPPSN 1200
 Db 1141 LPTKAVQPAVCVIGMYDTAQNDELAFSGQIINVLNKDPPMWKGEVSGQVGLPPSN 1200

QY 1201 YVKLTITDMDPSQO 1213
 Db 1201 YVKLTITDMDPSQO 1213

RESULT 2
ID ITN1 RAT STANDARD: PRT: 1217 AA.
AC 09WVE9; 09WVE1.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Intersectin 1 (EH domain and SH3 domain regulator of endocytosis 1).
GN ITSN1 OR ITSN OR EHSN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=9303609; PubMed=10373452;
RA Okamoto M., Schoch S., Suedhof T.C.;
RT "HSH1/Intersectin, a protein that contains EH and SH3 domains and
RT binds to dynamin and SNAP-25. A protein connection between exocytosis
RT and endocytosis?";
RL J. Biol. Chem. 274:18446-18454(1999).
CC -1- FUNCTION: Adapter protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters
CC several dynamin in a manner that is regulated by alternative
CC splicing. Also binds clathrin-associated proteins and other
CC components of the endocytic machinery, such as N-WASP, Eps15 and
CC Stomatin 2 (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC Enriched in synaposomes.
CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms: 1 (shown here) and 2;
CC are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
CC -1- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,
CC bind to dynamin.
CC -1- DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: ARI27798; AAD30271.1; -;
DR EMBL: ARI32672; AAD31026.1; -;
DR HSSP: P29354; IGFC.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 5.
DR Pfam: PF00016; ehfand; 3.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 5.
DR SMART: SM00054; Efh; 2.
DR SMART: SM00027; EH; 2.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS50031; EH; 2.
DR PROSITE: PS50002; SH3; 5.
KM Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
KM Alternative splicing.
FT CA BIND 21 109 EH 1.
FT CA BIND 66 178 EF-HAND 1 (POTENTIAL).
FT DOMAIN 221 310 EH 2.
FT DOMAIN

FT CA BIND 267 279 EF-HAND 2 (POTENTIAL).
FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
FT DOMAIN 350 670 COILED COIL (POTENTIAL).
FT DOMAIN 737 803 SH3 1.
FT DOMAIN 910 968 SH3 2.
FT DOMAIN 999 1057 SH3 3.
FT DOMAIN 1071 1135 SH3 4.
FT DOMAIN 1152 1211 SH3 5.
FT DOMAIN 321 324 POLY-SER.
FT VARSPLIC 1003 1073 MISSING (IN ISOFORM 2).
SQ SEQUENCE 1217 AA; 137154 MW; 6C13238A5E5B34B CRC64;

Query Match 96.0%; Score 6021; DB 1; Length 1217;
Best Local Similarity 95.7%; Pred. No. 3,3e-228;
Matches 1166; Conservative 19; Mismatches 27; Indels 6; Gaps 2;

QY 1 MAQFPFRGSLIDVAIVTEERAKHDQPLSLKPLAGITGDDQARNFFGSGLPQPVLAQ 60
DB 1 MAQFPFRGSLIDVAIVTEERAKHDQFOSLKPISGFTGDDQARNFFGSGLPQPVLAQ 60
QY 61 IVALDMNNDGMDQVERSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAFGIGI 120
DB 61 IVALDMNNDGMDQVERSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAFGIGI 120
QY 121 ASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLPANGAPVYIQPLPAFPAATWP 180
DB 121 ASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLPANGAPVYIQPLPAFPAATWP 180
QY 121 ACPMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLPANGAPVYIQPLPAFPAATWP 180
DB 121 ACPMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLPANGAPVYIQPLPAFPAATWP 180
QY 181 KSSSFRSGSGSLQNTLTKQKASFDVASAPAAEVAWVQSSRLKTRQLFNSHDKTMSHL 240
DB 181 KSSSFRSGSGSLQNTLTKQKASFDVASAPAAEVAWVQSSRLKTRQLFNSHDKTMSHL 240
QY 241 TGPQARTIMQSLPQAOASIMWLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 300
DB 241 TGPQARTIMQSLPQAOASIMWLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 300
QY 301 PEYIPSPFRVSSGSGSVSSVDQRLPEBSSSEDDQPEKKLPVTFEDKKNFERG 360
DB 301 PEYIPSPFRVSSGSGSVSSVDQRLPEBSSSEDDQPEKKLPVTFEDKKNFERG 360
QY 361 SVELKRRALLLEQQRKQERLAQLEAEQERKEREQERKQLELEKQERLELER 420
DB 361 SVELKRRALLLEQQRKQERLAQLEAEQERKEREQERKQLELEKQERLELER 420
QY 421 QEEERKKEIEREAKKELEERQLEMERNRQBLNQRKKEQEGTVLAAKRTLEFE 480
DB 421 QEEERKKEIEREAKKELEERQLEMERNRQBLNQRKKEQEGTVLAAKRTLEFE 480
QY 481 LEALNDKQHOLEGKLQDRCRLATQROEIESTNKSRELRIAEITHLQOQLQESQMLGRL 540
DB 481 LEALNDKQHOLEGKLQDRCRLATQROEIESTNKSRELRIAEITHLQOQLQESQMLGRL 540
QY 541 IPEKQILSDQLKOVQNSLHRDSLTLTKALPAKLANQQLREQLDEVEKTRSLQLEID 600
DB 541 IPEKQILSDQLKOVQNSLHRDSLTLTKALPAKLANQQLREQLDEVEKTRSLQLEID 600
QY 601 VNNQKLEIREIHSKQQLQKORSLEAERLKQKEERKSLLEKQKEBQORVOERDQWL 660
DB 601 VNNQKLEIREIHSKQQLQKORSLEAERLKQKEERKSLLEKQKEBQORVOERDQWL 660
QY 661 EHVQOEQPRPKPHEEDRLKREDSVRKKEAEERAKPEMDQKOSRLLFPHQEPKLAQ 720
DB 661 EHVQOEQPRPKPHEEDRLKREDSVRKKEAEERAKPEMDQKOSRLLFPHQEPKLAQ 720
QY 721 PMSTTEKGPLTISAQESAKVYVYRALYFPESRSHDEITIQGDIVM-----VDSQTEP 775
DB 721 PMSTTEKGPLTISAQESAKVYVYRALYFPESRSHDEITIQGDIVM-----VDSQTEP 775
QY 776 GWLGEELGKGTWFPANAAEKIPENEVPTPAKPYTDLISAAPKLAARETAPAPVTSSE 835
DB 776 GWLGEELGKGTWFPANAAEKIPENEVPTPAKPYTDLISAAPKLAARETAPAPVTSSE 835
QY 836 PSTTNNMADESSITWSSSNEKEPETDNDWDTAAQPSLTVPASGLRORSAPFPAATGSS 895
DB 836 PSTTNNMADESSITWSSSNEKEPETDNDWDTAAQPSLTVPASGLRORSAPFPAATGSS 895

```

Db      840 PSTPNWMAFSSSTWSPSTNEKPEPTDWDMTWAQPSLTVPASQDLRQSAFTPATATGSS 899
      896 PSPVLGGGEKEVEGIQAAALYPMRAKKNNHINPKSDVITLLEQCDMMWPEEVOGQGMFP 955
      900 PSPVLGGGEKEVEGIQAAALYPMRAKKNNHINPKSDVITLLEQCDMMWPEEVOGQGMFP 959
      956 KSYVKLTSGPVRKSTSDTGTPTESPAKLKVASPAKPAIPEGEEFIAMTYEESSEGDLLT 1015
      960 KSYVKLTSGPVRKSTSDTGTPTESPAKLKVASPAKPAIPEGEEFIAMTYEESSEGDLLT 1019
      1016 FQGGDIVVTYKQDGMWTVGVGDKSVFPPSNVYRLKDSGSGTAGTSGIAGKPEIAQVY 1075
      1020 FQGGDIVVTYKQDGMWTVGVGDKSVFPPSNVYRLKDSGSGTAGTSGIAGKPEIAQVY 1079
      1076 ASYAATGPEOULTLAPGQLILIRKKNPGWMEGELQARCKKROIGWPFANYKLLSPGTSK 1135
      1080 ASYAATGPEOULTLAPGQLILIRKKNPGWMEGELQARCKKROIGWPFANYKLLSPGTSK 1139
      1136 ITPELEKRTAVQAPVCCVIGMYDYTAQNDELAFSKQIINVINKEDPDWKGEGVSGQVG 1195
      1140 ITPELEKRTAVQAPVCCVIGMYDYTAQNDELAFSKQIINVINKEDPDWKGEGVSGQVG 1199
      1196 LFPSNYVKLTITDMPDSQ 1213
      1200 LFPSNYVKLTITDMPDSQ 1217

RESULT 3
ITN1_HUMAN STANDARD; PRT: 1721 AA.
ID ITN1_HUMAN
DC Q15811; 095216; 090NK2; 090U92; 090UET5; 090UK1; 090UK60;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Intersectin 1 (SH3 domain-containing protein 1A) (SH3P17).
CN ITSN1 OR ITSN OR SH3DLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Fetal brain;
RX MEDLINE=99017974; PubMed=9799604;
RA Gaipponi M., Scott H.S., Chen H., Schebesta A., Rossier C.,
RA Antonarakis S.E.;
RT "Two isoforms of a human intersectin (ITSN) protein are produced by
RT brain-specific alternative splicing in a stop codon.";
RL Genomics 53:369-376(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=99415290; PubMed=10482960;
RA Pucharcos C., Fuentes J.-D., Casas C., de la Luna S., Alcantara S.,
RA Adonès M.L., Soriano E., Estivill X., Pritchard M.;
RT "Alternative cloning of human intersectin (ITSN), a putative
RT multivalent binding protein expressed in proliferating and
RT differentiating neurons and overexpressed in Down syndrome.";
RL Eur. J. Hum. Genet. 7:704-712(1999).
RN [3]
RP SEQUENCE OF 963-1721 FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RA Teyba L.O., Kvaesa S.M., Skripkina I.Y., Anoplienko O.V., Slavov D.,
RA Tassone F., Rynditch A.V., Gardiner K.;
RT "House homologs of human chromosome 21 genes.";
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 620-1721 FROM N.A. (ISOFORM 3).
RC TISSUE=Bone marrow;
RX MEDLINE=98294438; PubMed=9630982;
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
RT "Cloning of ligand targets: systematic isolation of SH3 domain-
RT containing proteins.";
```

```

RL Nat. Biotechnol. 14:741-744(1996).
RN [5]
RP ALTERNATIVE SPLICING.
RC TISSUE=Fetal liver; and Brain;
RX MEDLINE=21548828; PubMed=11690630;
RA Pucharcos C., Casas C., Nadal M., Estivill X., de la Luna S.;
RT "The human intersectin genes and their spliced variants are
RT differentially expressed.";
RL Biochim. Biophys. Acta 1521:1-11(2001).
RN [6]
RP GENE MAPPING.
RX MEDLINE=98127038; PubMed=9465890;
RA Chen H., Antonarakis S.E.;
RL "The SH3DLA gene maps to human chromosome 21q22.1--q22.2.";
CC CYTOGENET. Cell Genet. 78:213-215(1997).
CC FUNCTION: Adapter protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles. Isoform 1
CC could be involved in brain-specific synaptic vesicle recycling.
CC SUBUNIT: Interacts with dynamin, SNAP-25, and SNAP-23. Clusters
CC several dynamin in a manner that is regulated by alternative
CC splicing. Also binds clathrin-associated proteins and other
CC components of the endocytic machinery, such as N-WASP, Eps15 and
CC Stomatin 2 (By similarity).
CC SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC Enriched in synaptosomes (By similarity).
CC ALTERNATIVE PRODUCTS: At least 4 isoforms; 1/long form/ITSN-1
CC (shown here), 2/short form/ITSN-s, 3/short form 2/SH3P17 and 4;
CC are produced by alternative splicing. Alternative splicing affects
CC domains involved in protein recognition and thus may play a role
CC in selecting specific interactions.
CC TISSUE SPECIFICITY: Ubiquitous in adult and fetal tissues,
CC except isoform 1 which is expressed almost exclusively in the
CC brain. Highly expressed in skeletal muscle, heart, spleen, ovary,
CC testis and all fetal tissues tested. Expressed at lower levels in
CC thymus, blood, lung, liver and pancreas. Isoform 1 is expressed in
CC all brain regions; not expressed in the spinal cord.
CC DOMAIN: SH3-3, SH3-4 and SH3-5 but not SH3-1 and SH3-2 domains,
CC bind to dynamin (By similarity).
CC DOMAIN: The KLRKQ domain binds to SNAP-25 and SNAP-23 (By
CC similarity).
CC DISEASE: Overexpressed in brain from Down syndrome foetuses
CC suggesting a gene dosage-dependent contribution to the
CC abnormalities of Down syndrome.
CC SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC SIMILARITY: CONTAINS 1 PH DOMAIN.
CC SIMILARITY: CONTAINS 5 SH3 DOMAINS.
CC SIMILARITY: CONTAINS 2 EH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@sib.ch).
CC -----
DR EMBL AF064244; AAC78611.1; -
DR EMBL AF064247; AAC80437.1; -
DR EMBL AF064245; AAC80437.1; JOINED.
DR EMBL AF064246; AAC80437.1; JOINED.
DR EMBL AF064243; AAC78610.1; -
DR EMBL AF114488; AAD29953.1; -
DR EMBL AF114487; AAD29952.1; -
DR EMBL AF180522; AAD53183.1; -
DR EMBL U61166; AAC50592.1; ALT_INIT.
DR HSSP; P29355; 1SEM.
DR Genew; HGNC:6183; ITSN1.
DR MIM; 602442; -
DR Interpro; IPR000008; C2.
DR Interpro; IPR002048; EF-hand.
```



```

DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF000018; SH3; 5.
DR Pfam: PF00036; ehand; 3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhGEF; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 4.
DR SMART: SM00326; SH3; 4.
DR PROSITE: PS0004; C2_DOMAIN 2; 1.
DR PROSITE: PS0010; DH 2; 1.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00031; EH; 2.
DR PROSITE: PS00003; PH_DOMAIN; 1.
DR PROSITE: PS00002; SH3; 5.
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
KW Alternative splicing.
FT DOMAIN 21 109 EH 1.
FT CA_BIND 66 77 EF_HAND 1 (POTENTIAL).
FT DOMAIN 221 310 EH 2.
FT CA_BIND 267 278 EF_HAND 2 (POTENTIAL).
FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KIERO).
FT DOMAIN 351 705 COILED COIL (POTENTIAL).
FT DOMAIN 740 806 SH3 1.
FT DOMAIN 913 971 SH3 2.
FT DOMAIN 1002 1060 SH3 3.
FT DOMAIN 1074 1138 SH3 4.
FT DOMAIN 1155 1214 SH3 5.
FT DOMAIN 1237 1423 DH.
FT DOMAIN 1462 1571 PH.
FT DOMAIN 1583 1679 C2_DOMAIN.
FT VARSPIC 770 MISSING (IN ISOFORM 3).
FT VARSPIC 1006 1076 MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT VARSPIC 1221 1721 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT CONFLICT 114 114 P -> A (IN REF. 2).
FT CONFLICT 1088 1088 T -> A (IN REF. 3).
FT CONFLICT 1109 1109 G -> R (IN REF. 3).
FT CONFLICT 1137 1137 E -> N (IN REF. 1).
FT CONFLICT 1361 1361 E -> A (IN REF. 2).
FT CONFLICT 1474 1474 N -> S (IN REF. 2).
SQ SEQUENCE 1721 AA; 195532 MW; 8189DDA0F145B4B5 CRC64;

Query Match 93.2%; Score 5842.5; DB 1; Length 1721;
Best Local Similarity 92.5%; Pred. No. 4.3e-221;
Matches 1128; Conservative 39; Mismatches 46; Indels 7; Gaps 3;

```

```

Qy 360 GSVLEKRRQALLLEQKREERLAQLERAEOERKEREROEAKRQLELEKORELE 419
Db 361 GNLELEKRRQALLLEQKREERLAQLERAEOERKEREROEAKRQLELEKORELE 420
Qy 420 ROREERREKTEIRRAARELEROROLEMERNRROELLNKNKEOEGVVLKARKTLEF 479
Db 421 ROREERREKTEIRRAARELEROROLEMERNRROELLNKNKEOEDVLKARKTLEF 480
Qy 480 ELEALNDKKHQLLEGKLDIIRCLATQROEISTYNSRELRAETTHLQQOQOESQOMLGR 539
Db 481 ELEALNDKKHQLLEGKLDIIRCLATQROEISTYNSRELRAETTHLQQOQOESQOMLGR 540
Qy 540 LIPEKQILSDQLKQVOQNSLRHDSLLTLKRALAEKELARQQLREQLDEVERETRSKLQEI 599
Db 541 LIPEKQILSDQLKQVOQNSLRHDSLLTLKRALAEKELARQQLREQLDEVERETRSKLQEI 600
Qy 600 DVFNQQLKELREIHSKQOLQORSLEAARLKQKQERSLELEKQKEDAQORVERBQKOW 659
Db 601 DFNQQLKELREIHSKQOLQORSLEAARLKQKQERSLELEKQKEDAQORVERBQKOW 660
Qy 660 LEHVQOE-EOPPRKPHBEDRLKREDSVRKKEAEERAKPEWODKOSRLFHPHOBPAKLAT 718
Db 661 LEHVQOEDEHORPKLHEERLKEEBSYKXKDGEBKQOEADKGLFPHQOBPAKPAV 720
Qy 719 QAPWSTTEKGLTISAQSVKVVYRALYPRESSHDEITIQPDIVV- - - -VDESQTG 773
Db 721 QAPWSTTEKGLTISAQSVKVVYRALYPRESSHDEITIQPDIVV- - - -VDESQTG 780
Qy 774 EPGWLIGELKGTGFPFANVAEKIPENEVPTPAKPYVDLTSAPAPKALARTPAPLPVS 833
Db 781 EPGWLIGELKGTGFPFANVAEKIPENEVPAKPYVDLTSAPAPKALARTPAPLVTS 840
Qy 834 SEPTTPNNMADFSTWSSSNEKPEDTNMDTMAOPSLTVPSAGOLRORSAFPTATG 893
Db 841 SEPTTPNNMADFSTWSTTNEKPEITNMDMAAQSLSLTPSAGOLRORSAFPTATG 900
Qy 894 SSPSPVLQGEKEVEGLQAQALYPWAKKKNLNFNKSQDVITLVEQDMMWGEVQOGKW 953
Db 901 SSPSPVLQGEKEVEGLQAQALYPWAKKKNLNFNKSQDVITLVEQDMMWGEVQOGKW 960
Qy 954 PPKSVYKLIISGPVRSSTSIDTGPESPASLKVASPAKPAIPBEETIANTYSSSGD 1013
Db 961 PPKSVYKLIISGIRKSTMSDSSSPASLKVASPAKPAIPBEETIANTYSSSGD 1020
Qy 1014 LTFQOGDVIIVTKKDGDMWTGTVGKSGVFPSPNVYRLKDSGSGTAGTSGLGKKPDLAQ 1073
Db 1021 LTFQOGDVIIVTKKDGDMWTGTVGKAGVFPSPNVYRLKDSGSGTAGTSGLGKKPDLAQ 1080
Qy 1074 VIASYAATGPQLTLAPQOLILIRKKNPGWMEGELQARGKKRQIGFPPANYVVLSPGT 1133
Db 1081 VIASYAATGPQLTLAPQOLILIRKKNPGWMEGELQARGKKRQIGFPPANYVVLSPGT 1140
Qy 1134 SKITTELPKTAQVAVQVIGMTDYTAQNDDELAFSGQIIVNLAKEDPMWKGVEVSGQ 1193
Db 1141 SKITTELPKSTALAAVQVIGMTDYTAQNDDELAFNKGQIIVNLAKEDPMWKGVEVSGQ 1200
Qy 1194 VGLPSPNVYKLTMDPSQ 1213
Db 1201 VGLPSPNVYKLTMDPSQ 1220

```

```

RESULT 4
ITN1_XENLA STANDARD; PRT; 1270 AA.
AC 0422877
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Intersectin 1.
CN ITSN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

```

OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Oocyte;
 RX MEDLINE=99030416; PubMed=9813051;
 RA Yamabhai M., Hoffman N.G., Hardison N.L., McPherson P.S.,
 RT Castagnoli L., Cesarini G., Kay B.K.;
 "Intersectin, a novel adaptor protein with two eps15 homology and five
 src homology 3 domains";
 RL J. Biol. Chem. 273:31401-31407(1998).
 CC -!- FUNCTION: Adaptor protein that may provide indirect link between
 the endocytic membrane traffic and the actin assembly machinery.
 CC May regulate the formation of clathrin-coated vesicles.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
 CC -!- BIRTHED IN synaposomes (By similarity).
 CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 EH DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF032118; AAC73068.1; -;
 DR HSSP; P29355; ISEM.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; EPS15_repeat.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 5.
 DR Pfam; PF00036; ehfand; 2.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00054; Efh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 2.
 DR PROSITE; PS50002; SH3; 5.
 KM Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding.
 FT DOMAIN 21 109
 FT CA BIND 66 77 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 220 309
 FT CA BIND 266 309 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 266 277 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
 FT DOMAIN 349 691 COILED COIL (POTENTIAL).
 FT DOMAIN 732 793 SH3 1.
 FT DOMAIN 897 955 SH3 2.
 FT DOMAIN 986 1044 SH3 3.
 FT DOMAIN 1058 1122 SH3 4.
 FT DOMAIN 1139 1198 SH3 5.
 SQ SEQUENCE 1270 AA; 143670 MW; EA940C1F6B6A858 CRC64;
 Query Match 80.8%; Score 5063.5; DB 1; Length 1270;
 Best Local Similarity 80.6%; Pred. No. 7.6e-191;
 Matches 980; Conservative 98; Mismatches 123; Indels 15; Gaps 8;

QY 179 WPKSSFSRSGSGSNTLTKLOKASFDVASAPPAEAWVPQSSRLTKYROLFNSHDKTMSG 238
 DB 178 LPKSSSFRSVAGSQINTKLOKASFDVAPPAEAWVPQSSRLTKYROLFNSHDKTMSG 237
 QY 239 HLTGPQARTIMOSSLPOAQLIWNLSIDIDQDKLTAEFEIILAMHLIDVAMSGQPLPPV 298
 DB 238 NLTGQARTIMOSSLPOAQLIWNLSIDIDQDKLTAEFEIILAMHLIDVAMSGQPLPPV 297
 QY 299 LPPEYIPSPFRVSGSGSVYSSSDVDRLPEPSSSEDEQPEPKLPVTFEDKKRNF 358
 DB 298 LPPEYIPSPFRVSGSGSVYSSSDVDRLPEPSSSEDEQPEPKLPVTFEDKKRNF 357
 QY 359 RGSVLEKRRQALLQOEKOEERLAQLEAEKEREKEREKOEAKROLEKOEKREL 418
 DB 358 RNLLEKRRQALLQOEKOEERLAQLEAEKEREKEREKOEAKROLEKOEKREL 417
 QY 419 ERQREERERKEIEREAKRELEERQOLEMERNRROELNQRNEQEGTVLKARRTLE 478
 DB 418 ERQREERERKEIEREAKRELEERQOLEMERNRROELNQRNEQEGTVLKARRTLE 477
 QY 479 FELEALNDKQOLEBKLODIDICRLATQOELESTNKSRELIATITLQOOLQSSQMLG 538
 DB 478 FELEALNDKQOLEBKLODIDICRLATQOELESTNKSRELIATITLQOOLQSSQMLG 537
 QY 539 RLPEKQILSDQLKQVQONSJHRSDILTLKRALBAKELARQOLEQDEVERETRSKLOE 598
 DB 538 RLPEKQILSDQLKQVQONSJHRSDILTLKRALBAKELARQOLEQDEVERETRSKLOE 597
 QY 599 IDVFNNQLEKELEIHSKOOLKORSLEAARLKOEKOEKSLLEKOEKEDARQOEKQ 658
 DB 598 IDVFNNQLEKELEIHSKOOLKORSLEAARLKOEKOEKSLLEKOEKEDARQOEKQ 657
 QY 659 WLEHYQOEOPRPRKPHEDRLKREDSYRKKEAEERAKPEMDKOSRLFHPHOEPAKLAT 718
 DB 658 WODRVKQEBE - RYKQOEKKEKKEESYQKCEVEK - KPELOEKPNKPFHPHOPKYLKG 713
 QY 719 QAPWSTTEKGPLTISAQESVKKVYVYRALLYPSSSHDEITTPQDVIWVDESQTEBQWL 778
 DB 714 QIPMNTEKAPLTIN - QGDVKKVYVYRALLYPFDARSHDEITTEPDIIMVDESQTEBQWL 772
 QY 779 GGELKGTGMPFANYAEKIPENEVPTPAKPYTDLTSAAPKALRETPAPL - PVTSSPS 837
 DB 773 GGELKGTGMPFANYAEKIPENEVPTPAKPYTDLTSAAPKALRETPAPL - PVTSSPS 828
 QY 838 TTPNNMADPSSSTWSSSNEKETDNDMTMAAOPSLTVPSAQOLRORSAFTPATATGSSPS 897
 DB 829 TNSNMADPSSSTWPTNNMTDKVESDNDMTMAAOPSLTVPSAQOLRORSAFTPATATGSSPS 888
 QY 898 PVLGQGEKVEGLOAQAALYPMRAKKDNHNLNFKSDVITVLEQDDMMWPGEVGQKGMFPKS 957
 DB 889 PVLGQGEKVEGLOAQAALYPMRAKKDNHNLNFKSDVITVLEQDDMMWPGEVGQKGMFPKS 948
 QY 958 YKLSISGPRKSTSIDTGPTESPALSKRVASPAKPAIIGEFTIAMTYESSSEGGDLTFQ 1017
 DB 949 YKLSISGPRKSTSIDTGPTESPALSKRVASPAKPAIIGEFTIAMTYESSSEGGDLTFQ 1008
 QY 1018 QGDVIVTKKDGDMWTGTVGDKSGVPSNYYRLKDSSEGSTAGTGSIGKPELIAQVIAS 1077
 DB 1009 QGDVIVTKKDGDMWTGTVGDKSGVPSNYYRLKDSSEGSTAGTGSIGKPELIAQVIAS 1068
 QY 1078 YAATPEQTLTAPGQILIRKKNPGWMEGELQARGKROIGWPPANVYKLISLSTGSKIT 1137
 DB 1069 YAATPEQTLTAPGQILIRKKNPGWMEGELQARGKROIGWPPANVYKLISLSTGSKIT 1128
 QY 1138 PTELPEKTAQVAPACOVIGYDYTAQNDELAFSKQOINVINKEDPDWKKKEVSGQVGLF 1197
 DB 1129 PTELPEKTAQVAPACOVIGYDYTAQNDELAFSKQOINVINKEDPDWKKKEVSGQVGLF 1188
 QY 1198 PSNYYKLITDMDPSQ 1213
 DB 1189 PSNYYKLITDMDPSQ 1204

RESULT 5
ID ITN2_HUMAN STANDARD; PRT; 1696 AA.
AC O9NZM3; O9NZG0; O9ULG4; O9NZM2; O9HAK4; O9NXB6; O9S062; Q15812;
AD 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Intersectin 2 (SH3 domain-containing protein 1B) (SH3P18) (SH3P18-like
WASP associated protein).
GN ITSN2 OR SH3D1B OR SWAP OR KIAA1256.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RC TISSUE=Fetal brain, and Fetal liver;
RX MEDLINE=20382722; PubMed=10922467;
RA Pucharcos C., Estivill X., de la Luna S.;
RT "Intersectin 2, a new multimodular protein involved in clathrin-
mediated endocytosis.";
RL FEBS Lett. 478:43-51(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Seifert M., Engel M., Weiler C.;
RT "Intersectin 2 (SH3D1B), human homolog of mouse Ees2 protein.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirotsawa M., Nomura N.,
Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [4]
RP SEQUENCE OF 258-720 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 1261-1696
FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
Tanase T., Nomura Y., Togiya S., Komai R., Hara R., Takeuchi K.,
Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 821-1284 FROM N.A. (ISOFORM 1).
RA Wilson L.A., Fields D., Cruz L., Friesen J., Siminovich K.A.;
RT "SH3P18-like wasp associated protein (SWAP): A multiple SH3 domain
containing protein that interacts with WASP.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 944-1191 FROM N.A. (ISOFORM 4).
RC TISSUE=Bone marrow;
RX MEDLINE=98294438; PubMed=9630982;
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowles D.M., Kay B.K.;
RT "Cloning of ligand targets: systematic isolation of SH3 domain-
containing proteins.";
RL Nat. Biotechnol. 14:741-744(1996).
RN [7]
RP ALTERNATIVE SPLICING.
RC TISSUE=Fetal liver, and Brain;
RX MEDLINE=21548828; PubMed=11690630;
RA Pucharcos C., Casas C., Nadal M., Estivill X., de la Luna S.;
RT "The human intersectin genes and their spliced variants are
differentially expressed.";
RL Biochim. Biophys. Acta 1521:1-11(2001).
CC -!- FUNCTION: Adapter protein that may provide indirect link between
the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.

CC -!- SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,
CC ITSN2 and Eps15, and different partners according to the step in
CC the endocytic process.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms; 1/ITSN2-L (shown here),
CC 2, 3/ITSN2-S1 and 4/ITSN2-S2/SH3P18; are produced by alternative
CC splicing.
CC -!- TISSUE SPECIFICITY: Ubiquitous. Isoform 1 is primarily expressed
CC in adult heart and liver.
CC -!- MISCELLANEOUS: Overexpression results in the inhibition of the
CC transferrin uptake and the blockage of the clathrin-mediated
CC endocytosis.
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF182198; AAF59903.1; -;
DR EMBL: AF182199; AAF59904.1; -;
DR EMBL: AF248540; AAF63600.1; ALT_INIT.
DR EMBL: AB033082; BAA86570.1; ALT_INIT.
DR EMBL: AK021545; BAB13841.1; -;
DR EMBL: AK000302; BAA91068.1; -;
DR EMBL: AF001630; AAD00899.1; -;
DR EMBL: U61167; AAC50593.1; -;
DR HSSP: P29355; 1SEM.
DR GeneW: HGNC:6184; ITSN2.
DR MIM: 604464; -;
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EFS15_repeat.
DR InterPro: IPR000261; EFS15_repeat.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 5.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhGEF; 1.
DR PRINTS: PR00499; P67PHOX.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM0054; Efh; 2.
DR SMART: SM0027; EH; 2.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhGEF; 1.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS50004; C2 DOMAIN_2; 1.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS50003; PH DOMAIN; 1.
DR PROSITE: PS50002; SH3; 5.
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
KW Alternative splicing.
FT CA_BIND 21..109
FT CA_BIND 66..78
FT DOMAIN 243..332
FT DOMAIN 360..755
FT DOMAIN 756..817
FT DOMAIN 897..955
FT DOMAIN 980..1038
FT DOMAIN 1052..1116
FT DOMAIN 1126..1185
EH 1.
EF_HAND (POTENTIAL).
EH 2.
COILED COIL (POTENTIAL).
SH3 1.
SH3 2.
SH3 3.
SH3 4.
SH3 5.

CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
 CC - SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.
 CC - SIMILARITY: CONTAINS 1 EF-HAND.
 CC - SIMILARITY: CONTAINS 5 SH3 DOMAINS.
 CC - SIMILARITY: CONTAINS 2 EH DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF132479; AAD19747.1; -
 DR EMBL: AF132480; AAD19748.1; -
 DR HSSP: P29355; 1SEM.
 DR MGD: MGI:1338049; SH341B.
 DR InterPro: IPR000008; C2.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000261; EPS15_repeat.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000219; RhGEF.
 DR Pfam: PF00018; SH3; 5.
 DR Pfam: PF00036; etHand; 2.
 DR Pfam: PF00168; C2; 1.
 DR Pfam: PF00169; PH; 1.
 DR Pfam: PF00621; RhGEF; 1.
 DR PRINTS: PRO0499; P67PHOX.
 DR PRINTS: PRO0452; SH3DOMAIN.
 DR SMART: SM00239; C2; 1.
 DR SMART: SM00054; EFh; 2.
 DR SMART: SM00027; EH; 2.
 DR SMART: SM00233; PH; 1.
 DR SMART: SM00325; RhGEF; 1.
 DR SMART: SM00326; SH3; 5.
 DR PROSITE: PS50004; C2_DOMAIN_2; 1.
 DR PROSITE: PS50010; DH_2; 1.
 DR PROSITE: PS00018; EF_HAND; 1.
 DR PROSITE: PS50031; EH; 2.
 DR PROSITE: PS50003; PH_DOMAIN; 1.
 DR PROSITE: PS50002; SH3; 5.
 DR Endocytosis: SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF-HAND (POTENTIAL).
 FT DOMAIN 244 333 EH 2.
 FT DOMAIN 364 716 COILED COIL (POTENTIAL).
 FT DOMAIN 717 778 SH3 1.
 FT DOMAIN 851 909 SH3 2.
 FT DOMAIN 941 999 SH3 3.
 FT DOMAIN 1013 1077 SH3 4.
 FT DOMAIN 1087 1146 SH3 5.
 FT DOMAIN 1169 1356 DH.
 FT DOMAIN 1395 1505 PH.
 FT DOMAIN 1517 1613 C2 DOMAIN.
 FT VARSPLIC 1187 1197 DPEQVIEVFQ -> GLQFEEQKTL (IN ISOFORM
 FT VARSPLIC 1198 1658 MISSING (IN ISOFORM 2).
 FT CONFLICT 451 452 KQ -> NT (IN REF. 1; AAD19748).
 FT CONFLICT 1161 1161 D -> G (IN REF. 1; AAD19748).
 SQ SEQUENCE 1658 AA; 188776 MW; 7050EFC5F7983A5 CRC64;

 Query Match 48.6%; Score 3049.5; DB 1; Length 1658;
 Best Local Similarity 52.4%; Pred. No. 3.5e-112;
 Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;

Db 61 IVALDLNKGDMQDEFSIAMKLKLGQYGLPSTLPPVWKQFV--AISAAPAFGIG 118
 119 GIASNP-----PLTAVAPV--PMG-----SIPVGVSPPLVSVPPAAVPLANGAP 164
 120 ---SNPNLSIHOPLPVAPVPLATPLSSATSGTISPLMPAPLVSSTSSLP---NGIAS 173
 165 VIQPLPAVAPPAATPKSSPS--RSGPSQNLTKLQKAS--PDVASAP----- 211
 174 LIQPL-SIPYSSSTLPHMSSVLMWGGF---GASIQAGQILIDGSSSSSTASLSGN 229
 212 ----AAEWAPQSSRLKXROLFNSHDKTMSGHLTGPOARTLMQSSIPQALASIMNLS 266
 230 SPKGTISEWAVPQPSRLKXROLFNSHDKTMSGHLTGPOARTLMQSSIPQALASIMNLS 289
 267 DIDQGLTAEFFILAMLIIVAMSGOPLPVLPPEVIPEPRFVRSSSGMSVSSSDVD 326
 290 DIDGQGLKAEFFILAMLIIVAMSGOPLPVLPPEVIPEPRFVRSSSGMSVSSSDVD 341
 327 QRLPEPSSSEDOPEKKLPVTFEDKKEENFERGSVELEKRRQALLQORKEORLAQLE 386
 342 GLTPSYQKTLQ-EEEPQKTLPTFEDKRRANERKGMELKRYLMQOKEAEARKQKE 400
 387 RAEQRKERERQOEAKQLELEKOLEKOLEKOLEKOLEKOLEKOLEKOLEKOLEKOLE 446
 401 KEWERKORELOEOWKQLELEKOLEKOLEKOLEKOLEKOLEKOLEKOLEKOLEKOLE 460
 447 EWERNRRELNORKEOEGVIVLKARKTLEFLEALNDKQHLLEGLOIRRLATQR 506
 461 EWERNRRELNORKEOEGVIVLKARKTLEFLEALNDKQHLLEGLOIRRLATQR 520
 507 QEISTNKSRELRLAETIHLQOQLOESQOMLGRILPEKQILSDOLKQOQNSLHRDILT 566
 521 TELEVLIDKQOLEIMEIYQLOQELKEYONKLIYVPEKQILNEIKMQLSNTDSSGL 580
 567 LKRALFAVELARQOLREOLDEVERETRSKLOEIVFNQNLKELEIHSKQOLQKORSLEA 626
 581 LHKSSSEKEELCORLEKQLEDALEKETAASKLEMDSPFNQLEKELSEYNTQALAEOLHKI 640
 627 ARKQKEERKSLLEKOK---EDAQRVQERDKOMLEHQOEOQRRPRRPHEDDLK 682
 641 KRDKLKEIERKRLQOLKQLEDEAKAKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 694
 683 EDSVAKKEAEERAKPEMDQKQSLFHPHOEPAKLATQAPWSTTEKGLTISAOESVAVV 742
 695 EKSQKQTEER---KAKAKQSE-----TASA-----LVN 721
 743 YRALYPPESRSHDEITQPGDIIVWVDESQGEPEGLGELKGTGMPFANYAEKI PENEV 802
 722 YRALYPPFARNHDEMSFSSGDIIOVDEKTVGPEGMVYGSFGKFGMPFANYAEKISSE 780
 803 PTPAPFVDTLSAPAPKALRETPAPLPVTSSEPTTPNNMADSSSTWPSNKKPETDN 862
 781 -----KALSPKALLPPTVLSLATSTS--SQPASVTDHNV--SESNLTVNT-- 824
 863 WDTAAQPSLITVPAGOLQRSATTPATATGSSSPVLAGQEGEVOAOLYPMRAKQD 922
 825 --TW-----QKSAFTRTVSPG-SVSPHGGQOAVENLKAQALCSWTAKKE 867
 923 NNLNFKSDVITVLEQODMMWFGVQOKGMPKSYVYKLSG-EVRRS--TSIDTGPTE 979
 868 NNLNFKSDVITVLEQODMMWFGVQOKGMPKSYVYKLSG-EVRRS--TSIDTGPTE 927
 980 PASLKRVSAPAKPAIP--GEEFIAMTYESESQODLTFQOQDVIVTKQDMMWTVGD 1038
 928 PTS---TAVPTSTRVAPVGEVYIALYSSVPEBDLFTFGEEELVTVQKGEWWTGSGE 984
 1039 KSGVFPNNYVRLKDSGSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGT 1098
 985 RTGIFPSVYVAPKQOENNGNASKSGANKKPEIAQVISAASAESTBDSLAPQGLIILK 1044
 1099 KNPQGWEGEGLQARGGKQKQIGWPPANYVYKLSPEKTSITITELPKTAVQPAVCVIGMD 1158

DB 1045 KNTSGWMOGELQARGKQKQWFPASHVKLLGPSSERTMPT-----FHAVQVIAMWD 1097
 QY 1159 YTAONDELAFSGQIINVLNKEDPDMKGEVSGQVGLFPSPNVYKLTMDPSQQ 1213
 DB 1098 YMANDELFNSGQILINWKNDDPPDMOGFTNGLTGLFPSNVYKMTTDSDSQQ 1152

RESULT 7
 EPI5_HUMAN
 ID EPI5_HUMAN STANDARD: PRT; 896 AA.
 AC P42566;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Epidermal growth factor receptor substrate 15 (Protein Eps15) (AF-1P protein).
 GN Eps15 OR APLP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RX MEDLINE=94239734; PubMed=8183552;
 RA Wong W.T., Kraus M.H., Carlomagno F., Zelano A., Druck T., Croce C.M., Huebner K., di Fiore P.P.;
 RA "The human eps15 gene, encoding a tyrosine kinase substrate, is conserved in evolution and maps to 1p31-p32.";
 RT Oncogene 9:1591-1597(1994).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94181254; PubMed=8134107;
 RA Bernard O.A., Mauchautte M., Mecucci C., van den Berghe H., Berger R.;
 RA "A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4, AF-9 nor ENL.";
 RT Oncogene 9:1039-1045(1994).
 RL
 CC - FUNCTION: INVOLVED IN CELL GROWTH REGULATION. MAY PLAY A ROLE IN SIGNAL TRANSDUCTION AND MITOGENICITY.
 CC - SUBUNIT: Interacts with Stonin 2.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC - TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
 CC - PTM: PHOSPHORYLATED ON TYROSINE BY EGFR.
 CC - DISEASE: INVOLVED IN A T(1;11)(P32;Q23) CHROMOSOMAL TRANSLOCATION IN ACUTE LEUKEMIAS CAUSING FUSION TO THE TRITHOMAS (MLL OR HRX) GENE PRODUCT WHICH CONTAINS DNA-BINDING MOTIFS RESULTING IN A ROGUE ACTIVATOR PROTEIN.
 CC - SIMILARITY: CONTAINS 3 EH DOMAINS.
 CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U07707; AA52101.1; -;
 DR EMBL; Z29064; CA82305.1; -;
 DR HSSP; P02649; 1B24.
 DR Genew; HGNC:3419; EPS15.
 DR MIM; 600051; -;
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_repeat.
 DR InterPro; IPR003903; UTM.
 DR Pfam; PF00036; ehand; 4.
 DR SMART; SM00054; EFn; 3.
 DR SMART; SM00027; EH; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 3.
 KW Phosphorylation; Calcium-binding; Repeat; Chromosomal translocation;

KW Proto-oncogene.
 FT DOMAIN 15 104 EH 1.
 FT DOMAIN 128 216 EH 2.
 FT CA BIND 173 184 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 224 314 EH 3.
 FT CA BIND 236 247 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 599 601 15 X 3 AA REPEATS OF D-P-F.
 FT REPEAT 623 625 1.
 FT REPEAT 629 631 2.
 FT REPEAT 634 636 3.
 FT REPEAT 640 642 4.
 FT REPEAT 645 647 5.
 FT REPEAT 651 653 6.
 FT REPEAT 651 653 7.
 FT REPEAT 664 666 8.
 FT REPEAT 672 674 9.
 FT REPEAT 692 694 10.
 FT REPEAT 709 711 11.
 FT REPEAT 737 739 12.
 FT REPEAT 798 800 13.
 FT REPEAT 804 806 14.
 FT REPEAT 825 827 15.
 FT DOMAIN 768 850
 FT MOD_RES 132 132
 FT PRO-RICH.
 FT PHOSPHORYLATION (BY TYR-KINASES)
 FT (POTENTIAL).
 FT M -> I (IN REF. 2).
 FT CONFLICT 822 822
 FT SEQUENCE 896 AA; 98673 MW; A1B9FB04A07FABEB CRC64;
 FT SQ

Query Match 7.8%; Score 486.5; DB 1; Length 896;
 Best Local Similarity 20.5%; Pred. No. 1; e-12;
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

QY 15 WAITVEERAKHDOQFLSPKPIAGFITDQARNFFQSGLPQVLAQIWMALDNNNGRMD 74
 DB 122 WAKVEDAKKDAFDLSIPVNGFLSGDKVKXPVLNKGKLPVILIGRWELSDIDHDGMLD 181
 QY 75 QVESIANKLIKIKLQGYQFSTLPVWKQOPVAISSAPAGIGIASMPPLTAVAPVM 134
 DB 182 RDEFVAMFLVYCALE
 QY 135 GSIPVGNSPPLSVPPPAVPPPLANGAPVIGPLPAFAPPAATWPKSSFSRSGPSOL 194
 DB 204
 QY 195 NTKLOKASPDVAPAPPAEWAVPQSSRLKYROLFNSHDKTMSGHLTGPARTLMOSL 254
 DB 216 KTV-----WVSPAEKAKYDEIFLKTDXMDQFVSGLEVEIREIFLKTGL 257
 QY 255 POQLASTIWNLSDDIDQDKLTAEEFILAMHLDVAM-SGQPLPPLPPEYTPSFRVRYS 313
 DB 258 PSTLTAHWSLCDTRKCGKLSKDPALAFHLISQKLKIGIDPPVILPPEMIPPS--DRA 314
 QY 314 GSGMSVSSSVQRLPREPESSEDOQPEKKLPYTFEDKKENFERGSVELEKRRQALL 373
 DB 315 SIQKNITGSSPV-----ADFSAIK-----ELDTLNNEIVD 344
 QY 374 QGRKEQRIALQERAEQRRKEREROEPAKQOLEKQLEKQRELEKQREERKEIERR 433
 DB 345 LQREK-----NNVEQDLKEKE-----DTIQRTSEVQ--DIQDEVQRE 380
 QY 434 EAKRELERORQLEWERNRROELNQRNKEQEGTVLKARRKLEFELALNDKQHQLEG 493
 DB 381 NTNLOKLQKQKQ-----QVQELDLE-----LDEQKQQLQEE 410
 QY 494 KLQDIRCLATQROEISTNKSRELRLAETHLQOQLOESQOMIGRLIPKQILSDQKQ 553
 DB 411 QLKEVRKKCAEBAQIISLQ-----AELTSQESQISTYEELAK-----AREELSR 456
 QY 554 VQONSIRHDSLTITKRALEAKELAQOQLRQQLDVEVERETR-----KLOEDIVNNQTK 607
 DB 457 LQOETALE-----LEESVESGAQLEPLQOHLQDSQDSQISNMQKLMEMKQLENHNSQLN 510
 QY 608 ELREIHS-----KOOLQKORSLAERLQKQBERKSIETL----- 641

QY 609 LREHSHKQOLQO-----RSLFARLKKQKQKRSLELEKQKQDQRPVQ 653
 Db 508 -----SNWSSSPSVLVNGATDYCSLSTSSSETANFNEHAEGNNLESETHOESSVRSS 562
 QY 654 ER-----DKOWLEHYOEOEPKPRKHEEDRLKREDSYRKKEA-----EERAKPEMODK 702
 Db 563 PELAPSDVTESEAVYVAAKGEKVTPR--FDODKHSKEDDFEVNESSSLTDAVDNLDFE 620
 QY 703 QSRIF-----HPHOBRKATLQAPNMTTEKGPLTISAQSVAVVYV 743
 Db 621 QSDPFVSGDPEFKDDPFKIDPFGGDPPKSGDPPASDCEFFQJSTDPPTTSTDPFASASN 680
 QY 744 RALYPFESRSHDEITIGDIIWVWDESQTEGP--GWLGLKLGKGTGWPANVAEKIPENE 801
 Db 681 SSNTSVETWKNHNDPFAAGGIVVAAASATDPPFASVFGNSFG--DGFADSTLSKXNNED 739
 QY 802 VPTP-----AKPVTDLTSA-----PA--PKLALRETPAPLIVTSSPSTTPNN 842
 Db 740 AFNPTISSSTSVIAKPMLEETAKSESDVPALPPKVGTPTRPCPP-----PGKRPIN 794
 QY 843 WADPSSST-----WPSSSNEKPEPTDNDWDMQAQPSLTIVSAQQLRQSAFTPATGSS 895
 Db 795 KLUSSDPLKLNDFPQPPGNDSPKXKDPDMFC-----DPTSTTTNKE 838
 QY 896 PSP 898
 Db 839 ADP 841

RESULT 9
 ID PANI YEAST STANDARD; PRT; 1480 AA.
 AC P32521;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE PANI protein.
 OS PANI OR MIP3 OR YIR006C OR YIB6C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxId=4932;
 RN
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 320-344; 352-375 AND 899-906.
 RX MEDLINE=92405166; PubMed=1339314;
 RA Sachs A.B., Deardorff J.A.;
 RT "translation initiation requires the PAB-dependent poly(A)
 RT ribonuclease in yeast";
 RL Cell 70:961-973(1992).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Uggels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odeh C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=95282515; PubMed=7762303;
 RA Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,
 RA Schwager C., Zimmermann J., Sander C., Anseorge W.;
 RT "Nucleotide sequence and analysis of the centromeric region of yeast
 RT chromosome IX";
 RL Yeast 11:61-78(1995).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.
 CC -1- SIMILARITY: CONTAINS 2 EH DOMAINS.
 CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A SUBUNIT OF PAB-DEPENDENT
 CC POLY(A)-SPECIFIC RIBONUCLEASE.

CC -----
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z38062; CA86208.1; -;
 CC EMBL; M90688; AAA34841.1; -;
 CC EMBL; X79743; CAB38097.1; -;
 CC PIR; S30889; S30889.
 CC PIR; S48440; S48440.
 CC SGD; S0001445; PANI.
 CC InterPro; IPR000261; BPS15_repeat.
 CC Pfam; PF00035; ehnd; 3.
 CC SMART; SM00054; Efn; 2.
 CC SMART; SM00027; EH; 2.
 CC PROSITE; PS50031; EH; 2.
 KM Repeat.
 FT DOMAIN 142 575
 FT REPEAT 142 153
 FT REPEAT 164 175
 FT REPEAT 188 199
 FT REPEAT 215 226
 FT REPEAT 235 246
 FT DOMAIN 270 359
 FT REPEAT 392 403
 FT REPEAT 409 420
 FT REPEAT 422 433
 FT REPEAT 446 457
 FT REPEAT 467 478
 FT REPEAT 498 509
 FT REPEAT 507 518
 FT REPEAT 538 549
 FT REPEAT 545 556
 FT REPEAT 564 575
 FT DOMAIN 328 680
 FT REPEAT 328 350
 FT DOMAIN 600 689
 FT REPEAT 600 689
 FT REPEAT 658 680
 FT DOMAIN 1084 1125
 FT REPEAT 1084 1089
 FT REPEAT 1090 1095
 FT REPEAT 1096 1101
 FT REPEAT 1102 1107
 FT REPEAT 1108 1113
 FT REPEAT 1114 1119
 FT REPEAT 1120 1125
 FT DOMAIN 1315 1377
 FT REPEAT 1315 1320
 FT REPEAT 1321 1326
 FT REPEAT 1327 1332
 FT REPEAT 1340 1345
 FT REPEAT 1345 1350
 FT REPEAT 1355 1360
 FT REPEAT 1361 1366
 FT REPEAT 1372 1377
 FT DOMAIN 13 22
 FT DOMAIN 29 34
 FT DOMAIN 98 106
 FT DOMAIN 1400 1406
 FT DOMAIN 1452 1455
 FT DOMAIN 1474 1480
 FT CONFLICT 235 235
 FT CONFLICT 266 273
 FT CONFLICT 474 487
 FT CONFLICT 553 557
 FT CONFLICT 1291 1291
 FT CONFLICT 1396 1480

15 X 12 AA APPROXIMATE REPEATS.
 1-1.
 1-2.
 1-3.
 1-4.
 1-5.
 EH 1.
 1-6.
 1-7.
 1-8.
 1-9.
 1-10.
 1-11.
 1-12.
 1-13.
 1-14.
 1-15.
 2 X 23 AA APPROXIMATE REPEATS.
 2-1.
 EH 2.
 2-2.
 7 X 6 AA TANDEM REPEATS.
 3-1.
 3-2.
 3-3.
 3-4.
 3-5.
 3-6.
 3-7.
 8 X 6 AA APPROXIMATE REPEATS.
 4-1.
 4-2.
 4-3.
 4-4.
 4-5.
 4-6.
 4-7.
 4-8.
 POLY-GLN.
 POLY-GLN.
 POLY-GLN.
 POLY-PRO.
 POLY-GLU.
 POLY-PRO.
 P -> T (IN REF. 1).
 ITAODAK -> YYCPKSGKN (IN REF. 1).
 MISSING (IN REF. 1).
 MISSING (IN REF. 1).
 A -> R (IN REF. 1).
 GSDVLPPEPLTQOASTSEPIAHVNYNGAEKGTAYGS
 DDDVLPPEPLTQOASTSEPIAHVNYNGAEKGTAYGS
 PPPLP -> EAFCLHPLHLYOLNKLPLONLSTLTITMVL

FT KKAHMDPILMTYRFLNOLVOMKRRKGNOLVOVSH
 SQ QFHLOVFLHPHFEDLICFL (IN REF. 1).
 SEQUENCE 1480 AA; 160267 MW; F3518495FF759553 CRC64;
 Query Match 6.5%; Score 410.5; DB 1; Length 1480;
 Best Local Similarity 20.5%; Pred. No.2,6e-09;
 Matches 261; Conservative 174; Mismatches 444; Indels 395; Gaps 47;
 17 ITVEERAKHDQOFLSLKPI-AGFTIGDQARNFFQSGLPQVPLAOTIWLADNNDGRMDQ 75
 266 ITAQDQAKFETLFRSIVTNGSNTVSGANCRLIMRSGLPQOLARIWLTCDTSKAGELLF 325
 76 VEFSTAMKILIKLQCYQPLSTLPPVMO----- 104
 326 PEFALAMHLINDVLOQDITLPEYLDSTKXNEVSSFIDAINLSIANODSSANDAKTPPDEF 385
 105 -----OPVALISSAPAFGI-----GGIASM----- 123
 386 ITAGVONLOPOTGYWPTQTSFGILPLOSITGGGVASALNPSTGEMAPPTFMKSMNTGTP 445
 124 ---PPLTAVAPVPM-----GSIP-VVGMSPPLVSVPPAAV-----PPLAN 160
 446 GINPOITGAPASMOININGNALQPTTGMPQTTGMPQTTGMPQTSFGVNLGPOLTG 505
 161 GAPPVQP-----LPAFAHPAATWPKSSFSRSGPQUL----- 194
 506 GA---LOSQYTGAGYGVMPQOSGPASM--PNLSFNQOGLQSLTGLQPTGFLPPSFS 560
 195 -----NTKLOKAQSFDAVASAPPAEMVAPQSRILKYQLFNHSHKTMGHLT 241
 561 ATMPPLTAQTTGGNNETLYKSNMNNNLINSSODKISTEESLFFYFIETPTONKGLD 620
 242 GPOARTILMOSSLPOQALASINWLSIDIDDGKLTAEFFILAMHLIDVAMSGPLPVP 301
 621 SPTAVIEFRKSGINRADLEQINWLCIDINNTGOLNKEFALGHVLYGKNGRPIINVLS 680
 302 EYIP-----PSF-----RVRVSG-- 314
 681 SLIPSTKLLDNLKNQKTEPTTTEKPSFGKIDALSYKNNDDVLPNYRNRKRYVSAKN 740
 315 ---SGMSVSSSVDRRLPEEPSSBEOQPEKKLPVTFEDK---KREFFEGSVLEKR 367
 741 EEOSSSSSSASAVNH-----SSSTLTQDDISVDKTVKKTAKRPYAGFSR---EINIK 791
 368 ROALLEOQKEOER-----LAOLERAEOEKEREQOEAQRQLE 407
 792 NIASLENEIKNISNPENCYDSSIPSDLTSRFDIAIAKLNLNEISTIDNETNANKIQLY 851
 408 LEKOLEKORELEOREERREKKEITERREAKRELEROROLEMERNNROELLNQRKEQGT 467
 852 RKK---NPSIIIGSGPNCGETENDRKKAKSRALLR-----ARMSALTGKSTESEDS 899
 468 VVLKARKKLEFE-LEALNDKHKQLEGKQDITRCRLATORQEIESTNKRRELRILEITHL 526
 900 LSNEDQOQSAEIKRIQOENKQOEI---IKDIRSSISDISASLKTMTGSMN---ISN- 951
 527 OQOLOQSOQMLGRLIPEKOILSDOLKOVOONSLHRS----- 563
 952 -QEFERWEGIGLEDVRRFL--DDLKSNKNSKVTSESPVPSTPTTPVDBRSSSPYSQF 1009
 564 LTLTKALBAKELIARQOLREQULDEVERETRSKLOEIDVFN--NQLKELREIHSKOOL-- 618
 1010 KYAEERAAALKEQAKKRMKELAKFPDKNRNVQSSRSISSENSREOQOIGASNVLVPR 1069
 619 ---OKQSLKARUKQOEKRSKLELEKOE----- 646
 1070 RATPFQEEKVEVAQPTQPVQSTQPVQPTQPVQPTQPVQPTQPVQPTQPVQPVQNY 1129
 647 DAQRAVOERDKQWLEHVQOEOEPKPKPHEDRLKR-----EDSVAKKEAEE----- 693
 1130 MAKQESDDEDEDEKRLQBELKRLKLTAKKADKEXKLLALRKQIDBAQNESSEEFINGKD 1189
 694 -----RAKP-----EMQDKOSRLFPHQEP--KLATQAPWSTTEKGLPTTISA 734

Db 1190 NEGVHNVPOAPAPAPSAFQNSVNAPRSVHAAVTPAACKNSTGLPSTMGHNPFRKDA 1249
 Qy 735 QESVYVYVYALYFFESRSHD-EITIQPDIVWVDESQTEPGMLGELKGTGWPPANY 793
 Db 1250 SAS-----STTFDRAAEMQRIORG-----LDEDE--DDQMSDEBENN--VAN 1292
 Qy 794 AEKIPENEVPTP---AKPYTDLTSAPAPKALARETPAP--LPVTSSEPTTPNNMADF 847
 Db 1293 DNKVEAKIGHPDHARAPV---TAAPLPSV-----TPVPVAVVPOANTSNKSSPIPIA 1345
 Qy 848 STWSSSNE-----KPETDWMDTMAQPSLTVPASQOLRQRS-AFTPATGSSSPVYG 901
 Db 1346 PIPPSVTQEPVPLAPLPVAVDGFQEPPIPSAPALATVQKSSGSTPALAGVLPVP--- 1402
 Qy 902 QGEVYEGLOAQALYPMRPAKKNLNFNK-----SDVITVLBCQODMMWFGVQOK 951
 Db 1403 --PPLPQOASTSPITIAHDYVNGAEKGTGAYGSDSDDVLSIPSS-----VGTBEBEE 1455
 Qy 952 GWPPKSYVKLISGP 965
 Db 1456 GAQPVSTAGIPISIP 1469
 RESULT 10
 YBET YEAST STANDARD; PRT; 1381 AA.
 ID YBET YEAST
 AC P34216;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 150.8 kDa protein in SEC17-QC1 intergenic region.
 GN YBL047C OR YBL0520 OR YBL0501.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE OF 1-961 FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95176707; PubMed=7871888;
 RA de Wergifosse P., Jacques B., Jomiaux J.-L., Purnelle B., Skala J.,
 RA Goffeau A.;
 RT "The sequence of a 22.4 kb DNA fragment from the left arm of yeast
 RT chromosome II reveals homologues to bacterial proline synthetase and
 RT murine alpha-adaptin, as well as a new permease and a DNA-binding
 RT protein.";
 RL Yeast 10:1489-1496(1994).
 RN [2]
 RP SEQUENCE OF 579-1381 FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=94205266; PubMed=8154187;
 RA Scherens B., el Bakoury M., Vierendeels F., Dubois E., Messenguy F.;
 RA "Sequencing and functional analysis of a 32,560 bp segment on the
 RT left arm of yeast chromosome II. Identification of 26 open reading
 RT frames, including the KIP1 and SEC17 genes.";
 RL Yeast 9:1355-1371(1993).
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOI/YBL047C FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 EH DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z35808; CA484867.1; -
 DR EMBL: X78214; CA455048.1; -
 DR EMBL: Z23261; CA480797.1; -
 DR PIR: S45781; S45781.
 DR SGD: S0000143; YBL047C.

DR InterPro: IPR000261; EPRS15_repeat.
 DR InterPro: IPR000449; UBA_domain.
 DR Pfam: PF00036; ehand; 2.
 DR Pfam: PF00627; UBA; 1.
 DR SMART: SM00054; EFH; 2.
 DR SMART: SM00027; EH; 3.
 DR SMART: SM00165; UBA; 1.
 DR PROSITE: PS50031; EH; 3.
 DR Hypothetical protein; Repeat.
 KW DOMAIN 14 113 EH 1.
 FT DOMAIN 135 227 EH 2.
 FT DOMAIN 277 366 EH 3.
 SQ SEQUENCE 1381 AA; 150783 MW; 626FDD61DCBA7D99 CRC64;

Query Match 6.4%; Score 403; DB 1; Length 1381;
 Best Local Similarity 20.0%; Pred. No. 4,7e-09;
 Matches 286; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

QY 3 QFPTPF-GGS-----LDWVAITVEERAKHDQFSLKPLAGITGDQARNFPFGSLP 54
 DB 109 QNPAPMOSGATGNTNNTDIPALSSNDIAKFSQLPDRTKACQVAGDKAKDIFLKARLP 168
 QY 55 QPVAQIATWALADNNNDGRMDQVEFSIAMKLKTLQGYOLPSTLPVYMKQOPVAISSAPA 114
 DB 169 NQTLGELIATLCDRDSAGVLDKSEFTIMAYLIQLCMSHHPSMNTPEAVLPQ----- 219
 QY 115 FGIGIASMPPLTAVAPVPMGSLPVVGMSPPLVSVPPAAVPLANGAPVYIOLPAPFAH 174
 DB 220 -----LWDSI-----RLEPVVAVNQPRTPPLSANST 245
 QY 175 PATWPKSSSFSGSGSGLNTKLOKQSPVAPAPAEAVAVQSSRLKYROLFNSHDK 234
 DB 246 GVSSLTRHSTISRLSTAFNSA-----ASDMSLSFEKKQGFDAIFDSDLK 230
 QY 235 TMSGHTLGPQARTILMOSSLPQAOLASTWNSIDIDQDKLTAEFFIAMLIDVANSQOP 294
 DB 291 QHAGSLSSAVLVPPFLSSRLNQETLATIWDLADIHNNAAEFTKLEFALAMELIQKNAGVE 350
 QY 295 LRPVLPPE-----YIPSPFRVRSQSGMSVSSSS----- 324
 DB 351 LRPVLPPELLOSALGLYPNPVLPQOQSAFOIALPSRSKPSLODMHGVSAVAVNTQPT 410
 QY 325 VDORLPPE-----EFS-----SEDEQOP 341
 DB 411 VPQVLPQNSNNSGLNDLLALNPSSSPPTKQTVVQNTNNSFYDNNNGQATLQOQOP 470
 QY 342 EKKLPVTFEDKKRENERGSGVELEKRRQALLQOKKEQERL----- 382
 DB 471 QQPPLTHSSSGLKPKFTPTS---NFGOSIIEKEPEEQQLRESSDTFSAQPPVPKHAAS 526
 QY 383 ----- 382
 DB 527 SPVKRTASTITLPQVPNFSVSPMPAGATSAATGAAGAAGAAALGASAFRSNNMAFKN 586
 QY 383 -----AQLEBAEQERKERERQEQEAKRQLELEKQLEKQERLEKQERERKEIER 432
 DB 587 QDLFADGEASQSLNATTEMANLSNQVNSLSKQASITTDKKSRAETOELKRYTEMKNSIQI 646
 QY 433 REAKRELERQOLEMENNRQOELLNQNKQEGVVLKARKLTLELELELANKKQOLE 492
 DB 647 KLNLRSTHDNVKQTEBLEAQVL---QVKNENE-----TLAQOLAVESENYIAAE 694
 QY 493 GKLDQIRCLATQROEISTYKSRRLIAEI---THLQOOLQESQOMLGRILPEKQILS 548
 DB 695 SKLME---LTTDQESQTKNAELKEQITNNSMTASIQSOLNKKQ----- 737
 QY 549 DQLKQVQONSJLHRSLLTLKALEKELARQOLREQLDEVERETRSKIQLEIDVNNQKE 608
 DB 738 -QVXQ-----ERSWVDVNSKOLEINQVTANLOKEIDGLG-----EKISVYLTKQKE 783
 QY 609 LREIHSKQOLQKORSLKARLQKQKE---RKSLEL---EKQKEDAQORVQCRDQWLEHV 663
 DB 784 LND-----YQKTVBEQHAQLOAKYQDLSKNDQTLTREKQLEERNRQIEQENLYHQHV 837

QY 664 QOEEQPRPKRHEEDRLKREDSYAKKEAEERAK-----PEMODKOSRLFHPHQP-- 713
 DB 838 SKIQE-----MFDLSQKXKSPFEKADQELKERNIEYANNVRESEKQMN--AMGQLEP 890
 QY 714 -----AKLATOAPWST---TEKGPL---TIS-----AQESKVVYVYALYPFESRSHDEI 757
 DB 891 AKDIITAKSASNTDTTKATKATSRGNVHEDTYSKFVETTVENSNLVNVKDEEKTERTES 950
 QY 758 TIQPGDI-VWVDESQGTGEPGLGELBKGTGWFPANYAEKIPF-----NEVLP-P-AKPV 809
 DB 951 DVFDVDPVPLTSGQSDSEBANTNNGQSGNETANP-NLTETLSDFRDDLNEYGI.PRSQSL 1009
 QY 810 TDLTSAPKPLALRETPAPLPVTSSEPTTNWADFSSTWSSSNKPEPTDNMDTWAAQ 869
 DB 1010 TSSVANNAPQ-SVRD-DVELPTELEERTT-NNTANRNDT---GNLSHPGEHEALPAT 1062
 QY 870 PSLTVPSAGOLRORSAPFPATATGSSSPVLG--QGEKVEGLQAO---ALVWRAKKN- 923
 DB 1063 ASTDVL-----NETTEVIEDGSTTKRANSNDEGESVSIQSPKISAQ-P-KAKTINE 1114
 QY 924 -----HLNFKSDVITYLLEQDMMWEGEVQOGKWPFKSVYKLSIGP----- 965
 DB 1115 EFPPIQELHIDESDSSDDDE---FEDTRE---IPSAVTKLQTPYNAQPTSSLEIH 1166
 QY 966 ---YKSTSIDPTSPESPASLKRVASPAKPAIP--GEFTIAMYTESSEQDGLTFQOGD 1020
 DB 1167 TEQVIRKYPAPGTPSPHNSGNSKASTNSILPVKDFDFEFAGL-----EQAAYEDNG- 1219
 QY 1021 VIVTKKDGDMWTGTGVGKSGVFPNSVYRLKDS-----GSGTAKGTSGLGKKEI 1071
 DB 1220 -----ADSESEFENVAVAGSMEQFETIDHLDLQMAFTGLTSSSNPTI-PRQV 1272
 QY 1072 AQVASYATGPEQLTLAPGLILIRKNPGGWMEGLQAKGKRQGWFPANYVKLISP 1131
 DB 1273 QO-----QSTSPAQV-----SNDEND-ELFA-----GF----- 1295
 QY 1132 GTSKITPTTEL--PKTAVQPAVCOVIGMYDYTAOND--ELAFSGQIIN 1176
 DB 1296 GNSKAEPTVATPSPQOP-----IPUKNDPIVDASLSKGPVYN 1334

RESULT 11
 TRHY RABIT
 ID TRHY RABIT STANDARD; PRT; 1407 AA.
 AC P37709;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trichohyalin.
 GN THH.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxId=9986;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Fietz M.J., Rogers G.E.;
 RL Submitted (DEC-1992) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAPFOLD PROTEIN, TOGETHER
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 CC DIFFERENTIATION.
 CC -!- SUBUNIT: HOMODIMER (PROBABLE).
 CC -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF

CC THE EPIDERMIS.
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 CC DIFFERENT SPECIES.
 CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z19092; CAA79519.1; -
 DR PIR: S28589; S28589.
 DR HSSP: P02633; 41CB.
 DR InterPro: IPR001751; CaBP S100.
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; efhand.1.
 DR Pfam: PF01023; S_100.1.
 DR ProDom: PD003407; CaBP_S100.1.
 DR PROSITE: PS00018; EF_HAND; 1.
 DR PROSITE: PS00303; S100_CaBP; 1.
 KM Keratinization; Repeat; Calcium-binding.
 FT DOMAIN 1 91 S-100 LIKE.
 FT CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
 FT CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
 SQ SEQUENCE 1407 AA; 183781 MW; A617D2A159F12B7F CRC64;

Query Match 6.2%; Score 391.5; DB 1; Length 1407;
 Best Local Similarity 29.7%; Pred. No. 1.3e-08;
 Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

QY 326 DQRLPEPSEDEQOEKLPVTFEDKKNFSGVLEKRRQALLLEQ----- 375
 DB 198 EEFIEEQLRRRQOEKRLREEREEQRRRRQHERALQEEBOLLRRMRREERQO 257
 QY 376 -----RKQERLALQERAEQERKERERQOEAKQOLEKQERQERQEEER 426
 DB 258 QLRRELEIRERQRLQERERREQRLQERERREQRLRELEIRERQERLEQEE 317
 QY 427 RKE-----IERE-AKREL-----ERQOLEERNRROELNORNEO---EETVLKAR 473
 DB 318 RRQRLEQERRRQRLQERLEIRERQERLEQEE-ERREQLAEVEEQAREKESITRW 376
 QY 474 RKLLEFELEALNDKQOLEKQODICRRLATQROEISTKSRLEIAITHTLOOQO-- 531
 DB 377 QRLQEEAGARQKQVSRPRQEEQSLRQOQERRO---RQERERLEEQARQOQOQWE 432
 QY 532 -ESQOMLGR-----IPEKOISDOLKVOO-----NSLHRSLLTKRALEKELARQO 580
 DB 433 EESERRRQRLARPSLRERQRLAEERQOEORPREEEQRRERROQLQEELEEQQRE 492
 QY 581 LRQQLQE-----VERERSKLOEI-----DVNNOLKE-----LRHIS-----KQOLQKRS 623
 DB 493 RAQQLQOEEDSFQERRERRRQOQORQEQTWQQLQEBQARRHTLYAKQOQOEQLREEE 552
 QY 624 LEARLQKEQKEKSLLEK--QKEDAQREVRQERDKQW--LEHVQOEOEPREPKEPDEDR 679
 DB 553 LQREKRRQ--ERREYREERELQREBDEKRRRQRRQRRQRYRLLELRLQEBQRLDKLREEQ 611
 QY 680 L---KREDSVRKKEAEERAKPEMQ 700

DB 612 LLQREBERLRQERERKLRKEEQ 635

RESULT 12
 TRHY HUMAN STANDARD; PRT; 1898 AA.
 AC Q07283;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trichohyalin.
 GN THH OR TRHY OR THL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 CX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93280194; PubMed=7685034;
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
 RA Steinert P.M.;
 RT "The structure of human trichohyalin. Potential multiple roles as a
 RT functional EF-hand-like calcium-binding protein, a cornified cell
 RT envelope precursor, and an intermediate filament-associated (cross-
 RT linking) protein.";
 RT J. Biol. Chem. 268:12164-12176(1993).
 [2]
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=93315697; PubMed=7686953;
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
 RA "Trichohyalin: a structural protein of hair, tongue, nail, and
 RT epidermis.";
 RT J. Invest. Dermatol. 101:65S-71S(1993).
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 CC DIFFERENTIATION.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 CC THE FLIPLOM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 CC THE EPIDERMIS.
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 CC DIFFERENT SPECIES.
 CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L09190; AAA65582.1; -

```

Db 571 REEKKRLQGEORRRQRKLKQGEERRDQLLKREERRQOQLKQEOERLEQRKLKREEVRLLE 630
Oy 581 LRQULDE-VERETRSKLQSDIVFNNOQLKELR---EHSKQOLQORSL---EAARLKOK 632
Db 631 QOEERDERLKRPEERERHERHLLKSEEOERHNEQLRREOQERKRQRLKREEBERLEQR 690
Oy 633 -----QOEKSLTLE-KQKEDAQORVQOERDKOMLEHNVQEOBPBRK----- 673
Db 691 LKREHEERERQGLABEEQOARERIKSRIPWQQLSEADARQSKYLEAPQAGRAA 750
Oy 674 PHEBRLKREDSYRKKEAEERAKPEMOKOSLFLPHPEPAKLATQAPWSTTEGPILTIS 733
Db 751 PQOEKERRRRESLQOEERAHARQOQOEERDFTWQOAE-----EKSEGRORLS 803
Oy 734 AQSVKVYVYRALYPESRSNDE 756
Db 804 APPLNRQRRQRRLRAEERQORQ 826

RESULT 13
TRHY SHEEP
ID TRHY SHEEP STANDARD; PRT; 1549 AA.
AC P22793;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trichohyalin.
GN THH.
OS Ovis aries (Sheep).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN (1)
RS SEQUENCE FROM N.A.
RX MEDLINE=93260018; PubMed=7684041;
RA Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;
RT "Analysis of the sheep trichohyalin gene: potential structural and
RL calcium-binding roles of trichohyalin in the hair follicle.";
RL J. Cell Biol. 121:855-865 (1993).
[2]
RS SEQUENCE OF 1016-1549 FROM N.A.
RP STRAIN=Merino-Dorset horn X Border Leicester; TISSUE=wool follicles;
RC MEDLINE=90130632; PubMed=2298812;
RX Fietz M.J., Presland R.B., Rogers G.E.;
RA "The CDN-deduced amino acid sequence for trichohyalin, a
RT differentiation marker in the hair follicle, contains a 23 amino acid
RL repeat.";
RL J. Cell Biol. 110:427-436 (1990).
[3]
RS FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLY).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE EPITHELIA OF THE TONGUE, HOOF AND RUMEN.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, COMPRISED A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.

```

CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 CC PROBABLY CONVERTED TO CITRULINES BY PEPTIDYLARGININE DEIMINASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z18361; CAA79165.1; -
 DR EMBL: X51695; CAA55992.1; -
 DR PIR: A34209; A34209.
 DR PIR: S32633; S32633.
 DR PIR: A40691; A40691.
 DR HSSP: P02633; 11G5.
 DR InterPro: IPR001751; CABP_S100.
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; efhand.1.
 DR Pfam: PF01023; S_100.1.
 DR ProDom: PD003407; CABP_S100; 1.
 DR PROSITE: PS00018; EF_HAND; 1.
 DR PROSITE: PS00303; S100_CABP; FALSE NEG.
 KW Keratinization; Repeat; Calcium-binding.
 FT DOMAIN 1 91 S-100 LIKE.
 FT CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
 FT DOMAIN 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
 FT REPEAT 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 449 476 1-2.
 FT REPEAT 477 504 1-3.
 FT REPEAT 505 532 1-4.
 FT REPEAT 533 560 1-5.
 FT REPEAT 561 588 1-6.
 FT REPEAT 589 616 1-7.
 FT REPEAT 617 644 1-8.
 FT REPEAT 645 678 1-9.
 FT REPEAT 679 706 1-10.
 FT REPEAT 707 742 1-11.
 FT REPEAT 743 771 1-12.
 FT REPEAT 772 796 1-13.
 FT REPEAT 797 832 1-14.
 FT DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 938 961 2-1.
 FT REPEAT 962 985 2-2.
 FT REPEAT 986 1021 2-3.
 FT REPEAT 1022 1044 2-4.
 FT REPEAT 1045 1067 2-5.
 FT REPEAT 1068 1090 2-6.
 FT REPEAT 1091 1121 2-7.
 FT REPEAT 1122 1144 2-8.
 FT REPEAT 1145 1167 2-9.
 FT REPEAT 1168 1197 2-10.
 FT REPEAT 1198 1227 2-11.
 FT REPEAT 1228 1250 2-12.
 FT REPEAT 1251 1273 2-13.
 FT REPEAT 1274 1296 2-14.
 FT REPEAT 1297 1319 2-15.
 FT REPEAT 1320 1342 2-16.
 FT REPEAT 1343 1368 2-17.
 FT REPEAT 1369 1391 2-18.
 FT REPEAT 1392 1416 2-19.
 FT REPEAT 1417 1439 2-20.
 FT REPEAT 1440 1461 2-21.
 FT REPEAT 1462 1484 2-22.
 FT REPEAT 1485 1507 2-23.
 FT VARIANT 1145 1197 MISSING (IN SHORT FORM).
 FT VARIANT 1251 1273 MISSING (IN SHORT FORM).
 FT CONFLICT 1399 E -> G (IN REF. 2).

SQ SEQUENCE 1549 AA; 201173 MW; E72F89FF1326E54E CRC64;
 Query Match 6.0%; Score 375.5; DB 1; Length 1549;
 Best Local Similarity 27.6%; Pred. No. 6,2e-08;
 Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;
 QY 332 EPSSDEQPEPKLPVTFE---DKRENFGSGVELKKRQALLBQKRQEPLAQLR 387
 Db 383 KPAQREQVREEDQRLKEKQLQREKRQREKQREYVELQREE--ERLQREEDQLQREER 440
 QY 388 AEOEKREERQ-----EQEAQQLLEKQ-----LEKRELERQREER 426
 Db 441 EKRRQREKQYLEKVELMEBQQLQREBKQREKQREYLEKVELBEBQQLQREBK 500
 QY 427 RKEIRREPAARLEQRLQREEMERNRQELQNKQEGTVLAKRKTLFELALND 486
 Db 501 RQERQRYLEKVELQEBEQLOREERK-----RQRE-----RQYLE----- 538
 QY 487 KKHQLEGLQDRIQRLATQROELSTNKSRELIAETTHLQ--QQLQSQQMLGRLLPEK 544
 Db 539 -KVELQEEQ-----LQREKREKQREKQYLEKVELQEBEQLOREBKQREK 590
 QY 545 QILSDQLKQVQNSLHRSLLTLKRALBAKELARQQLREQLDVERETRSKLQIDVFNN 604
 Db 591 QYL-EKVELQEEBQLQREBKQREKQREYLEKVELQEB--EQVQREK-----RR 641
 QY 605 QUKELREIHSKQQLQKQSLQEAARLKQEKQKSLQLEK-----QKEDAQK 650
 Db 642 QERREQYLEKQLQEBQLQEBEQLEBKREKQREKQREYLEKVELQEBEQLOREBK 701
 QY 651 RVQERDKQWL--EHVQEEQPRPKPHEDRLKQSDSVRKKEAER--AKPNQDKQSL 706
 Db 702 RQERERQYLEKEQLQREB--RLQREKQLOREDEKQREKQRYERKYLEBEEQLQEBDRL 758
 RESULT 14
 CALD_HUMAN STANDARD; PRT; 793 AA.
 ID CALD_HUMAN
 AC Q05682; Q13979; Q13978; Q14741; Q14742;
 DT 01-NOV-1997 (Ref. 35, Created)
 DT 01-NOV-1997 (Ref. 35, Last sequence update)
 DT 15-JUN-2002 (Ref. 41, Last annotation update)
 DE Caldesmon (CDM).
 GN CALDI OR CDM OR CAD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Lung fibroblast;
 RX MEDLINE=91358497; PubMed=1885618;
 RA NOVY R.E., Lin J.L.-C., Lin J.J.-C.;
 RT "Characterization of cDNA clones encoding a human fibroblast
 RT caldesmon isoform and analysis of caldesmon expression in normal and
 RT transformed cells";
 RL J. Biol. Chem. 266:16917-16924(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Aorta;
 RX MEDLINE=9220999; PubMed=1555769;
 RA Humphrey M.B., Herrera-Sosa H., Gonzalez G., Lee R., Bryan J.;
 RT "Cloning of cDNAs encoding human caldesmons";
 RL Gene 112:197-204(1992).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 5).
 RX MEDLINE=93101679; PubMed=1465449;
 RA Hayashi K., Yano H., Hashida T., Takeuchi R., Takeda O., Asada K.,
 RA Takahashi E.-I., Kato I., Sobue K.;
 RT "Genomic structure of the human caldesmon gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:12122-12126(1992).
 CC -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
 CC REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND

OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 RN [1] TaxID=4896;
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtz S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jaegels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean K.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Ocell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellern J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,
 RA Gabel C., Ruchs M., Fitzer C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe".
 RL Nature 415:871-880(2002).
 CC -1- SIMILARITY: CONTAINS 2 EH DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z69368; CA93290.1; -
 CC EMBL: Z70691; CA94638.1; -
 CC DR InterPro: IPR002048; EF-hand.
 CC DR InterPro: IPR00261; EFS15_repeat.
 CC DR InterPro: IPR003124; WH2.
 CC DR Pfam: PF00036; efhand; 3.
 CC DR Pfam: PF02205; WH2; 1.
 CC DR SMART: SM00027; EH; 2.
 CC DR SMART: SM00246; WH2; 1.
 CC DR PROSITE: PS50031; EH; 2.
 CC DR Hypothetical protein; Repeat.
 CC KW Hypothetical protein; Repeat.
 CC FT DOMAIN 281 369 EH 1.
 CC FT DOMAIN 821 910 EH 2.
 CC SEQUENCE 1794 AA; 193279 MW; 1F042BA5E805E27 CRC64;
 QY
 QY Query Match 5.2%; Score 327; DB 1; Length 1794;
 QY Best Local Similarity 20.9%; Pred. No. 5,6e-06;
 QY Matches 237; Conservative 137; Mismatches 404; Indels 356; Gaps 52;
 Db 676 QRTGQMPMPQRTGMPQ--MPGQGPMPAPQRTGMPMPQRTGMPQ-- 721
 QY 28 QPSTLRPIAFITGDARNFPPFQSLGPVLQAIMALADP--NNDRMQVERSIMKLI 85
 Db 676 QRTGQMPMPQRTGMPQ--MPGQGPMPAPQRTGMPMPQRTGMPQ-- 721
 QY 86 KXKLGQYQLPSTLPVWKQGPVAVISSAPAFGIGIASMPPLTIVAVPMSPITPV- 140
 Db 722 --MPGQGPMPA-PQRTGMPMPA-----PQRTGMPQMPG-PMLQRTG 761
 QY 141 GMSBPLVSVPPAAMP-----PLANGAPVITQPLPAFAPAAITWPKSSSPRSRSGPS 192

Db 762 GMAP-----QPTGMPGMPGINTPLSN--LPGIALGQGMMPNA-----PSG 801
 QY 193 QUNTLOKAQSFVVASAPPAEMAVAPQSSRLKRYQLPFSHDKTSGHLTGQARTILMQS 252
 Db 802 GINTTFQ--QKKDI-----PMASKEEKRIYDQIFDAMDKERKGTIGNAVLEIRGS 852
 QY 253 SLPLQALASIMWLSIDIDQKLTAEFLAMHLIDVAMSGPLPVPVPEVYIPSPFRVR 312
 Db 853 KLTRELEHINWLCDHGKGSLDRDEFVALHLILKRLNGEVAVLPPELIPSTRTFT 912
 QY 313 SSGSGSVSSSVQDRLPEPSSSEDEQPEKLVYTEDKRENFERSV--ELEKRRQ 369
 Db 913 --ESLNQVKNILKNDTSNRKPFGAENOSKLRK--NSFDNPSETTEKDATLYRHNDSDAS 968
 QY 370 ALLEQREOEERLAEQER 423
 Db 969 AVSSARRDRFKEEIESAPPIINDIDELISLKRKH-EKSLVNALEDKLLATPAND.1027
 QY 424 EERRKEIRREAPAKRELEROLEWERNRROELNORKEDEGTVLKARKTLEFELIA 483
 Db 1028 VQNDSLIRIKSVQDEINR-----LSTNKSPEVA-----SNMVRLEE 1065
 QY 484 LNDKRGLEGLQDTRCLATQROEISTNKSRLRI--AEIT--HLQQLQESQMLGR 539
 Db 1066 LSTRVSKM--LSDI-----NEVDHTISSLKLFQADRTNSYDQTSPEATOERNR 1114
 QY 540 LIPEQQLSDQKQVQNSLHRDSLILTKRLALEKELARQOLQROLEVERETSKY--- 596
 Db 1115 TISSK--LAEMEKQNES-----KALE-----OMKYVYNNIENNIKALIPS 1155
 QY 597 -----OEIDVFNQULKEIRHSKQOLQKORSLA---ARLK- 630
 Db 1156 AANDAWLSQWVDESVRVYVKELPVPAPAPQTLNPSPSVTQSKPIBENTHTPEVKA 1215
 QY 631 QKEQKSLLEKQ---KEDQRRVQERDKQLEHYQOEQPPPKPHEDRLKREDSV 686
 Db 1216 TSESPASNSNEDRAIRKAEORRMERLALGIRKOKTSPSPA-----PV 1263
 QY 687 RKKEEERKAKEMQKQ-----SRLFP-----HOEPKATLQAPW 722
 Db 1264 NSATSTPVAAFTAQIDQKASAVSSVNPVAVSISTPVAVPTVQHPOFTK--QIP- 1319
 QY 723 STTEKPLTISAQESVKKVYRYRALYPPESR----- 752
 Db 1320 TAAVADPSTTSTSTFTAPTPQA--PLENQFSKMSLEPPVPAVPTSPKQIPDSSNVHA 1377
 QY 753 -----SHDEITIQGDIWMDESQTEPGMLGELKGTGFPANVAEK---- 796
 Db 1378 PPPVQPMNAMPASHNAVAPSAPEERRDSFGSVSSGNSVSIIDETSTMPLKASQPTNPG 1437
 QY 797 IPENEV-----TPAKPVTLD--LTSAPAPKTLARTPTAPLVTSTSEPTTNMW 843
 Db 1438 APSNAPQVVPAPMAHVAAPQAPGKMTNAPAPSSA---PAPPAVSQLPAPVFN-- 1491
 QY 844 ADFSSTWSSSNEKRETDNMWTAQPSLTVPSAGQLRQSAFAPATATGSSPSVVGQ 903
 Db 1492 VVPEMISVVAQPPSS---VAAPATAPSTLTPS-----QSSF---AHVSPAPAPQH 1539
 QY 904 EKVGLQALQYPMRAKDNHLNFKNSKSDVITVLEQQDMWFGVEQKGMPKSVYKLIS 963
 Db 1540 PSAALSS-----APADNMPHRSRP-----YAPQF----- 1565
 QY 964 GPKRKSTIDT-----GPTF---SPASLRKVASPA--AKPAIP 996
 Db 1566 -PVQKPAIINNIPATNLGTSQSPSPKMGVNNNGSLPAMNAGQPSLAVPAV 1618

Search completed: March 14, 2003, 12:08:12
 Job time : 47 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:19:11 ; Search time 40 Seconds
(without alignments)
6248.379 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269
Sequence: 1 MAQFPTPGSLDVAITVE.....VGLFSPSYVKLTDDMPSSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeal: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1735	27.7	1011	5	O61639
2	1702	27.1	1097	5	O9VIF7
3	1698.5	27.1	1094	5	O61618
4	1333	21.3	1097	5	O9U2T9
5	794.5	12.7	662	5	O8T068
6	536.5	8.6	662	4	O8UBC2
7	528	8.4	1253	5	O8W61
8	527	8.4	907	11	O6O902
9	506.5	8.1	751	5	O23418
10	506.5	8.1	796	5	O9B1F4
11	493.5	7.9	599	11	O91WH8
12	485	7.7	668	5	O9W11
13	434.5	6.9	1116	3	O9HGL2
14	405	6.5	944	5	O18137
15	405	6.5	1148	5	O952Y7
16	405	6.5	1480	5	O18138

17	402.5	6.4	1003	12	O91LX9	O91LX9 kaposi's sa
18	397	6.3	1089	12	O40947	O40947 kaposi's sa
19	386	6.2	1036	12	O9DUM3	O9DUM3 kaposi's sa
20	386	6.2	2376	5	O966V1	O966V1 drosophila
21	383.5	6.1	1909	5	O25893	O25893 plasmodium
22	382	6.1	1162	12	O98148	O98148 kaposi's sa
23	377.5	6.0	1129	12	O9QR71	O9QR71 kaposi's sa
24	375	6.0	1151	4	O9BX65	O9BX65 homo sapien
25	367	5.9	976	12	O9DUN0	O9DUN0 kaposi's sa
26	363.5	5.8	2701	4	O9Y520	O9Y520 homo sapien
27	356	5.7	1743	5	O966V0	O966V0 drosophila
28	350.5	5.6	1108	5	O9ND10	O9ND10 babesia big
29	350	5.6	1750	3	O8X0H2	O8X0H2 neosporea
30	344.5	5.5	1004	4	O9UNH7	O9UNH7 homo sapien
31	344	5.5	330	11	O35146	O35146 mus musculu
32	344	5.5	420	11	O08635	O08635 caenorhabdi
33	342	5.5	2722	5	O19135	O19135 drosophila
34	341.5	5.4	1287	5	O9NGS5	O9NGS5 dicystoele
35	340.5	5.4	800	5	O9Y102	O9Y102 drosophila
36	336.5	5.4	1292	4	O9BX66	O9BX66 homo sapien
37	332.5	5.3	1639	5	O9VNF2	O9VNF2 drosophila
38	331.5	5.3	1069	4	O75137	O75137 homo sapien
39	331.5	5.3	1100	4	O9NV82	O9NV82 homo sapien
40	330	5.3	964	13	O42489	O42489 gallus gall
41	329.5	5.3	585	11	O9UK01	O9UK01 rattus norv
42	329.5	5.3	665	11	O92509	O92509 rattus norv
43	328.5	5.2	554	5	O9VPS3	O9VPS3 drosophila
44	326.5	5.2	665	11	O8R551	O8R551 mus musculu
45	324.5	5.2	1181	10	O9X116	O9X116 arabidopsis

ALIGNMENTS

RESULT 1

ID O61639 PRELIMINARY; PRT; 1011 AA.

AC O61639;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Dynamn associated protein isoform DAP160-2.
GN DAP160 OR CG1099.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R; TISSUE=HEAD;
RA Roos J, Kelly R.B;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
DR EMBL; AF054612; AAC39139.1; -.
DR HSSP; P29354; IGRI.
DR FLYBase; FBgn0023388; DAP160.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; ERS15_repeat.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; EFhand; 2.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 4.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS00018; EF HAND; UNKNOWN_2.
DR PROSITE; PS50002; SH3; 4.
KW SH3 domain.
SQ SEQUENCE 1011 AA; 111619 MW; 54B8589C92021C91 CRC64;

Query Match 27.7%; Score 1735; DB 5; Length 1011;

DR HSSP; P29354; 1GRI.
 DR FlyBase; FBgn023388; Dap160.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_repeat.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF00018; SH3; 4.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 4.
 DR SMART; SM00054; Eph; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 4.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00002; SH3; 4.
 DR Calcium-binding; SH3 domain.
 KW SEQUENCE 1097 AA; 120766 MW; 2573D7F8DCBACB CRC64;
 Query Match 27.1%; Score 1702; DB 5; Length 1097;
 Best Local Similarity 31.7%; Pred. No. 5, 6e-72;
 Matches 418; Conservative 203; Mismatches 329; Indels 368; Gaps 33;
 QY 11 SLVMAITVERAKHQDQFLSKPIAGFITGQDARNFPPQSGLPQVLAQIMALADMDND 70
 DB 4 AVDAVAATTPPERLKYOEQFALQPOAGFTGAAQKGFLLQSLPPLILQIMALADTSD 63
 QY 71 GRADOVEFSIAMKLKILKLOGLYLPSTLPVPMKQOFAISSAPAFGIGIASMP---LT 127
 DB 64 GKNINNEFSJACKLINLKGMDVPKVLPSTL-----LSSL-----TGDPVSMTPRGST 113
 QY 128 AVAPVPMGSIPIVVGMEPLVSVPPAAVPPPLANGAPPVIOPLPAFAHPAATWPKSSFSR 187
 DB 114 SLSPPLD---PLKGI--VPAAVAPVAPVAPVPA--VATVISP--PGVSPSGPPTPTSN-- 162
 QY 188 SGGSQQLNTLQKQAFSDVASAPPAEMAVPQSSRLKYRQLFNSHKTKMSGHLTGQART 247
 DB 163 --PPSRHTSISERAPSTESVN---QGEWAQQAQKRYTOVFANRDTRSGYLTGQANG 217
 QY 248 ILNQSSILPQALASIMWLSIDIDGKLTAEFLAMHLIDVAMSGOLPPLVLPPEYIPPS 307
 DB 218 VIVQSKLPQVTLAQIWLTDIDGRLNCFILAMFLCKMAAGKIPVTLQGEVWPN 277
 QY 308 FRRVRSGSGMSVISSSSVDQRLPEEPSESD-----EQQPEKKIP--VTEEDK 353
 DB 278 LRKIKSRPG---SVSGVSPRPGSPASRHASVSSGSGVVDADPTAG.PGQTSFEDKR 333
 QY 354 REMFERGSALEKRRQALLQEQKKEORLAQLERABERKERERKQEQAEKOLEKOLE 413
 DB 334 KENYVVKQALDRRRKIMEDQQRERERERERERERERERERERERERERERERER 393
 QY 414 KORELEKOREERERKETEREAKRELEPOROLEMERNRROELLNQNKQOEGTVLKAR 473
 DB 394 KOREIEMEKEQKRELEAKRELEKORQOEWEGARLAEMNAQKEREQOEVVLQOKH 453
 QY 474 RKTLEFELALNDKQOLEGKLODIRCLATOROEIESTNKSRELAETHTLQOOLQES 533
 DB 454 NTQJNELSTLNKIKELSORICDTRAGTVNVTVIDMKRTQDTSMSSEMSQAKRIKQ 513
 QY 534 QQNLGRILPK-----QILSDQLQVOQNSLHRSLTLTKALEAKTLAQOLR 582
 DB 514 NAKLLQTLQORAKWEAKSAGALGGENAQOQ-----LNAAFHKKOLIINOIR 563
 QY 583 EQLDEVERETRSKLOEIDVFNNOLEKLE-----IHSKQOLQORSLEAARLKQEQER 637
 DB 564 DKVENIKETLESKEDINTNDVQMSLKALSLALITCELYKEYIDVQRTSVLELKYNK 623
 QY 638 SLELEKQEDAQRRVQDRQWLEHVQOEOPRRKPEHEDRLKREDSVYKKEAERAP 697
 DB 624 NETSVSSAMPT-----GSSSAM-----EETGTVTDY-----AVASNDISALAAR 664
 QY 698 EMQDKQSRLEHPOEPKALQTAQWSTTEGPLTISQESVYVYVYVYVYVYVYVYVY 757
 DB 665 AVD-----LGGPAP-----EG-----FVKYQAVVEFNAARAEI 693

QY 758 TIQPDIVMDESQTEGPGWLGSLKGTGKGFNPANYAEKIPENVEVPIPAKPVYDILTSAPA 817
 DB 694 TFPQDILIVLEQNAEPCWLAGINGHTGWFPPESVYEKLEVEGV-----A 739
 QY 818 PKLARETPAPLPVTSSEPTTPNNMADFSSSTWSSSENKPEPTNNMTMAAQPSLTVPSA 877
 DB 740 PVAAVE--APVDAQVAVADTVN-----DNIT--SSIPAA 771
 QY 878 GQLRQSAFTPATATSSPSFVLQGEKEVEGLQALYPMWRKKDNHJNFKSDVITVLE 937
 DB 772 S-----ADLTAAQ-----DV----- 781
 QY 938 QQDMWFGEVQGGKWPSPKSYVKLISGPVRKSTISIDTGPTEPASLKRVASPAKPAIPG 997
 DB 782 ----- 781
 QY 998 EEFIAMTYESSBOGDLTFQGGDIVYTKKQDMDWMTGTGDKSGVPPSNVYRLKD----- 1052
 DB 782 EYIAAYPEESAERQDLSFSGAGEMVYIKKEGWMVTGIGRTQMPSPSNVYQKADVGTAS 841
 QY 1053 -----SRSQ----- 1057
 DB 842 TAAAPVESLDOETTLNQNAAVTAAPVDAQOYVQPLVQGEPSQPISSPGVGAEEAHED 901
 QY 1058 -----TAGKT-----GSLGKKPEIAQVIYASYATGPQOLT 1087
 DB 902 LDTEVSQINTQSTQSSSEPAESYGRPMSTRTSMTPGMRKAKSEIAQVIAPYEATSTQLS 961
 QY 1088 LAPQGLILIRKKNPGWMEGLQARKKRQIGWFPANVYVLLSG--TSKTPTELKTA 1145
 DB 962 LTRGQLIRKKTDSGWMEGLQAKRRRQIGWFPATYKVLQGGRRNSGRTPVSGSRIE 1021
 QY 1146 VQPAVC--QVIGMYDTAQNDELAFSKQOILINLVNKEGDPWMKGEVGOVGLFPSNVY 1202
 DB 1022 MTEQLDLKVIATLYPKAQNDELSPDKDILISVGRQBPWMKRELNGSLFPSNVY 1079
 RESULT 3
 061618 PRELIMINARY; PRT; 1094 AA.
 AC 061618;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Dyanin associated protein isoform DAP160-1.
 CN DAP160 OR CG1099.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1] TaxID=7227;
 RC SEQUENCE FROM N.A.
 RP TISSUE=HEAD;
 RA Roos J., Kelly R.B.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
 DR EMBL; AF053957; AAC9138.1; .
 DR HSSP; P29354; 1GRI.
 DR FlyBase; FBgn023388; Dap160.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001452; Eps15_repeat.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF00018; SH3; 4.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 4.
 DR SMART; SM00054; Eph; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 4.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
 DR PROSITE; PS00002; SH3; 4.
 KW SH3 domain.

SQ SEQUENCE 1094 AA; 120495 MW; 141370EB1FB960F0 CRC64;
 Query Match 27.1%; Score 1698.5; DB 5; Length 1094;
 Best Local Similarity 31.6%; Pred. No. 8,2e-72;
 Matches 417; Conservative 201; Mismatches 329; Indels 371; Gaps 33;

QY 11 SLDVVAITEERAKHQOFLSLKPIAGFITGDQARNFFPFGSLDPQVLAQIMALADNND 70
 DB 4 AVDMAVAVPRERLKYQEOFPALQFOAGFVTGAQAKGFLOQLPPLILGQIMALADTDSD 63
 QY 71 GRMOVSESIAMKLIKXLOGVLPSTLPPVMKQPVASISAPAFGIGIASMP---LT 127
 DB 64 GKMINNESIACKLNLKGMVFKVLPSSL-----LSL---TGDVPSMTPEGSTS 113
 QY 128 AVAPVPMGSIPIVGMSPPLVSVPPAAPPVPLANGAPVLOPFAFAHPAAATWPKSSFSR 187
 DB 114 SLSPLD---PLKGI-VPAVAVPVVPAVPAV---VATVISP-PGVSPSGTPTPTSN--- 162
 QY 188 SGPSQNTKIQAKQSPFVASAPPAEWAIVPQSSRLKXROLFNSHDKTMSGHLTGPAQT 247
 DB 163 ---PSRHTSISERAPSLIESVN---QGEWAQAAQRKXTQVFNNDRSGYLTGSGQARG 217
 QY 248 ILMGSLPQAOIASIWNLSIDIDQKLTAEFFILAMHLIDVAMSGOPLPVPPEYIPPS 307
 DB 218 VLVGSKLPQVTLAQIMTLSDIDGRLNCDFILAMFLCEKAMAGKIPVTLPOEWYPPN 277
 QY 308 FRVAVSGSGMSVSSSVDOFLPEPSSD-----EQOPEKLP--VTPEDKK 353
 DB 278 LRKTKSRG---SVSGVAVSRPGQAPASRNHVASVSGGVADDPAGJPGQTSFEDKR 333
 QY 354 RENFERGVELLEKRRALLEGQKOEERLAQERAEQEREREQOEQAEAROLELEQLE 413
 DB 334 KENVYKGAELDRRKRMEDQOKKERERERERERERERERERERERERERERERERER 393
 QY 414 KORELERQREERERKEIERREPAKRELERQLEMERNRROELLNQNKQOEGTVLAKAR 473
 DB 394 ROREIEMEKER 453
 QY 474 RKTLEFEALNDKKHOLEGLQDIPRCALQROEISTNSKRLRLAETHLOOQOES 533
 DB 454 NTQANVELSTINEKIKELSORICDTRAGVTNVKTVIDQMTQRTDSMSESQLAKRIEQ 513
 QY 534 QOMLGRLLPER-----QILSDQLQVQONSILHRDSLTLTKALEKEARQOLR 582
 DB 514 NAKLQLQLEBAKWEAKSKASGALGGENAQEQ-----LNAAPAHKOIINQIK 563
 QY 583 EQLDEVERETSQLOEIDVFNNOULKELE-----IHSKQLOKQRSLEAPARKQOEQERK 637
 DB 564 DKVENISKEIKKEDINTDVQMSLEKALSAITKCEDLYKEYDYQRTSVLELKYNRK 623
 QY 638 SLELEKQKEDAKRQVQERDKQWLEHVQOEQPRPKPHEBDRLEKREDSVKKAEERAKP 697
 DB 624 NETSVSSAMDT-----GSSSAM-----EETGTTVTDY-----AVASNDLSALAAP 664
 QY 698 EMODKQSKLFFHOEPAPKALQAPWSTTEKQPLTISAQSVKVVYALYPFESRSHDEI 757
 DB 665 AVD-----LGGPAP-----EG-----FKYQAVVEFNARNAEII 693
 QY 758 TIQCGDIYVWESQTEGFGWLGELKGTGTFPANYAKETIPENVPPAPKPYVDLTJAPA 817
 DB 694 TFFVGDIILVLEQNAEPEGWLAGELNGHTGTFPESYVEKLEVGEV-----A 739
 QY 818 PKLARETPAPLPVTSSEPTTPNNWADFSSSTWSSSNKEKETNMOTWMAQSPFLVPSA 877
 DB 740 PVAIV-EPAPVQAQVADT-----YNNINIT-----SSTIPA 768
 QY 878 GOLRORSAPFTATYTGSSPSVULQOGEKVBGLQAQALYPMWAKKONHLNFKSDVITVLE 937
 DB 769 S-----ADLTAAG-----DV----- 778
 QY 938 QQDMMWFEVGGQGWPFKSYVVKLISGVKRSSTIDGPTESPASLKRVASPAKPAIPG 997
 DB 779 ----- 778

QY 998 EEFIAMTYESESQDLTFQOQDVIVTKDQDMWTGTGVDKSGVSPSNVYRLD----- 1052
 DB 779 EYVIAVPPESAEEDLFFSAGEMVMVKKGEWMWTGTIGRTGMPFENYVQKADVGTAS 838
 QY 1053 -----SESSG----- 1057
 DB 839 TAAEPVESLDQETTLNGNAAYTAAPVEAOQVYQPLFVQEPSSOPISSPVGAEEAHED 898
 QY 1058 -----TAKT-----GSLGKPEIAQVIASYAATPEQLT 1087
 DB 899 LQTEVSQINTQKTSSEPPAESYRPMRTSSMTGMAKSELSIAQVIAPEATSTBLS 958
 QY 1088 LAPGOLLIRKKNPGGWMGEGELQARQKROIGWPPANYVLLSPG--TSKITPTELPKTA 1145
 DB 959 LTRGOLIMRKKTDSGWMGEGELQARQKROIGWPPANYVLLSPG--TSKITPTELPKTA 1018
 QY 1146 VQPAVC-QVIGWYDYTAQNDDELAFSGQIINVLNKEDPMMWKGVSQVGLFPSNVY 1202
 DB 1019 MTEQILDKVIALYPYKAQNDDELSPDKDITISVIGRDEPEMMWKGELNSGLFPSNVY 1076

RESULT 4
 Q9U2T9 PRELIMINARY; PRT: 1097 AA.
 AC Q9U2T9
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Y16A8C.36 protein.
 GN Y16A8C.36.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodetidae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA McMurtry A.A.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 investigating biology";
 RL Science 282:2012-2018(1998).
 CC -i- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
 DR EMBL; AL117204; CAB55138.1; -.
 DR HSSP; P29355; 1SEM.
 DR InterPro; IPR000194; ATPase a/bcentre.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR00261; EPS15_repeat.
 DR InterPro; IPR00108; Neu_cyt_fac2_2.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00459; P67PHOX.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS0018; EF_HAND; UNKNOWN_2.
 DR PROSITE; PS0002; SH3; 5.
 KW SH3 domain.
 SQ SEQUENCE 1097 AA; 122072 MW; 1C2BA5F103968372 CRC64;

Query Match 21.3%; Score 1333; DB 5; Length 1097;
 Best Local Similarity 29.4%; Pred. No. 1.1e-54;
 Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;

QY 15 MATVIERAKHQOFLSLKPIAGFITGDO-----ANFFQSGLPQVLAQIMALADN 67
 DB 15 MATVIERAKHQOFLSLKPIAGFITGDO-----ANFFQSGLPQVLAQIMALADN 67

```

Db 5 WEVSDAEYCKNFMF-----GOLTCGPFMAVTAARNALMRSNLPVOVLSQIMASDLS 57
Qy 68 NNDGRMDQVFFSIAMKLIKLIKLOGLYOSTLLPVMQOQVVAISSAPAFGIGIASMPPLT 127
Db 58 DKDGRDLIREYSIAMRLALNCLAGIPPIPOLPSSLVTPA--RNAPPTMGs----- 107
Qy 128 AVAPVPMGSLPVVGMSPVL-----SSVPPAVPPLANAPVIOPLPAFMAHAPATWIK 181
Db 108 -----RHGSVDYSQTLPPALIDRMSSQSYIPSA--PVSIASTP----- 142
Qy 182 SSSFSRSGSSQSLNTKLQKQASFDVASAPPAEMAVPQSSRLKYROLFNHSDKTMGHLT 241
Db 143 SSSRNSISAGSPLN-----NDRNVEFGROLENNWALPHNKKLYSOLFNALDERGSGIS 196
Qy 242 GPQARILMOSLPPQOASIMWLSDIDODKLTAEFFILAMHLIDVANSQGPLPVLP 301
Db 197 SQVGRSLGSLGPTVNLAMIFLSDVNMKGKLSVEYALISQWIMFSGFALPITL 256
Qy 302 EYIPPSFRVRSSGSGSVSSSVDRORLPPESSDEBOQPEKLPVTPEDKKEENFERGS 361
Db 257 ELV-----RMCGISSRSANNTPELEPGAEPPQ--KSPAPYTFEDKRODNLKSGQ 303
Qy 362 VELKRRQALLEOORKEORLAOLERAEOERERQEOEAKROLELEKOLEKORLEBO 421
Db 304 ABLERRRRVLEEEORRAAEVEKKEEREEAKKOREROEKROAEVEROELEORLIEBO 363
Qy 422 REE-----RKREIRR-----EAKRELE-----ROROLEMERNRROELLORNKOEQTV 469
Db 364 REEBEKKRRLEEMERREDEKMKRVOMEKAKVOMQVOMEQKKMFPYNOQOENRMLQ 423
Qy 470 LKARRKLELEALNDKQKQLEBKLODIRCRATOROEIESTNKSRELIETLHQO 529
Db 424 ROOREKTLQLOALDEKVIDEVDIGKAKEAVEVTGFIERRSTRDEKVARI-----KE 479
Qy 530 LQSSQMLGRLIEKOLISDOLKOVOONSILHRSLSLTLKALAKELARQOUREQD----- 586
Db 480 LQETNO-----KTAIESOLGHQOL--LOKQSAHKEI--TORKSELEALRRKRAIKALIEDNA 533
Qy 587 -EYERETRSLKLOEIDVFN--NNOLKELREIHSKQOLOKORSLAARLKQEOERKSLLEKO 644
Db 534 LELSTEREKSYNQTLEIKTKKEKYKTIVYSK-----LVAKREBYRNSFPL----- 578
Qy 645 KEDAOORRVOERDKOMLEHVOOEOPRPRKP-----HEEDRLKREDSVRK 688
Db 579 -----LVHAQTHARSKIGEFPAKSAAPASAPAPAPARTTTNGFPANFPAFGEFD--K 631
Qy 689 KEAEERAKPEMOKOSLFLPHROEPKAKLTQAPWSTTEKPEPLTISAOESKVVYVYALP 748
Db 632 TDASQRFDAFGATST-----ADPFAQIAQAP--AHSKGAVDOSAFNIDHYKCRALPA 683
Qy 749 FEARSHEITIOFGDIYVWDESQTEGPGWLGELKGTGFWPANYAEKIPENEVPTPAKP 808
Db 684 FEARSEBELSFEEDVDIVFOSHAEFGMAGOLREKVGFPFAFVBAIA--AVLPFG-- 739
Qy 809 VTDLTSPAKKALARETAPLPVYTSSEPTTPNNMADFSSSTWSSSNEKEPTDNMTWAA 868
Db 740 -----GDPRIQNMPEMNTPESSVDQIGV--KAARKEI-----AA 772
Qy 869 QPSLITVPSAQLOKRSAPFTPATATGSSPVLGOGKEVBLQOALYPMWAKKDNHLEN 928
Db 773 AMGLTEGGA-----PPASSAPAAVAIVIS-----QCIAGOFWARNEBELSLFA 814
Qy 929 KSDVITVLEQOQDMWFG--EVOGQKGMFPKSYVVLISGPRKSTSTIGTPESPASLSKRAV 987
Db 815 KSGDTIEVLEQEKMKGRNAPAGELGMFPKSYVKEVANTSTTPTI--VSPKASAGAPGAA 873
Qy 988 SPAAPAIPEE-----FIAMYTYSSEQDULTFQOGDVIVVTKKQDGMWMTG 1034
Db 874 AGAQYDVAVPDLTLOASETAPOOOLYVIVDFEAVETTDLALHVGDITLLEKNDWMWG 933
Qy 1035 TVGDKSGVFPNSYVRLKDSGSGTAGKT--GSLGKPEI--AQVLAASYATGPEDULTA 1089
Db 934 RCNGREGIFPANYVEISVQAGDPTPTPAAPPTVLCFAKVVDVFAVAPNOLGIK 993

```

```

Qy 1090 PGQILILINKNPGWBGSELQARGKROIGWFPANYVYKLSPTGSKITPTPELEKTAVOPA 1149
Db 994 VGEIVKIREKSAAGWEGELLIRNGKP--IAGWFPGEYVYVLEEAASPAT----- 1040
Qy 1150 VQOIVGMQDYTAQNDDLEAFSGQIINVLNKEDEPDMWG--EVSQGVGFPSNYVK 1203
Db 1041 -RATAVDYDEASQDDELGFKTGDTIVTDSKSEAMWSGHRBQPSKSGLFPNSYVO 1095

RESULT 5
Q8T068
ID Q8T068 PRELIMINARY; PRT: 662 AA.
AC Q8T068;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE LD23686P.
GN DAPI60.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.B., Rubin G.M., Celinker S.,
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069517; AAL39662.1;
SQ SEQUENCE 662 AA; 72553 MW; 3532CB7034BD7AC CRC64;

Query Match 12.7%; Score 794.5; DB 5; Length 662;
Best Local Similarity 25.9%; Pred. No. 1e-29;
Matches 221; Conservative 108; Mismatches 203; Indels 321; Gaps 21;

Qy 459 QRNKQEGTVLKKARKLFELEALNDKQKQLEBKLODIRCRATOROEIESTNKSREL 518
Db 4 QEROERVLKQKANHNTQJLVNELSTLNEKIKELSORICDTRAGTAVNTKTVIDGKRTQRT 63
Qy 519 RAEFTTHLOOQLOESQMLGRLIEK-----QILSDOLKOVOONSILHRSLSLT 567
Db 64 SMSMSQLKAKIKEONAKLTLOTOERAKTEAKSAAGALGEMAOBO-----L 113
Qy 568 KRALBAKELAQOLREOLDEVERETRSLKLOEIDVFNNOLEKLE-----IHSKQOLQOR 622
Db 114 NAAFAHQOLIINQIKDKVENISKEIESKEDINTNDVOMSELKALSLITKCEDLYKEY 173
Qy 623 SLEAARLKQEOERKSLLEBKQEDAOORRVOERDKOMLEHVOOEOPRPRRPHNEBCLKR 682
Db 174 DVORTSVLELKYNRKNETSVSAMD-----GSSASW-----EETGTVTVDPY----- 216
Qy 683 BDSVYKKEABERAKEMQDQSLFLPHROEPKAKLTQAPWSTTEKPEPLTISAOESVKVY 742
Db 217 -AVASNDISALABAVD-----LGGPAP-----EG-----FVK 243
Qy 743 YRALYFESRSHDEITIOFGDIYVWDESQTEGPGWLGELKGTGFWPANYAEKIPENEV 802
Db 244 YQAVEYFARNABEITFVPGDIILVPLEQNAEPGLMGEINGHTGWFPESEYVEKLEVGCV 303
Qy 803 PTPAPVTDLTSAAPKALARETAPLPVYTSSEPTTPNNMADFSSSTWSSSNEKEPTDN 862
Db 304 -----APVAAVE--APVDAQVATVADTYN-----DN 327
Qy 863 WDTWAQPSLTVPSAQLOKRSAPFTPATATGSSPVLGOGKEVBLQOALYPMWAKKD 922
Db 328 INT-----SSTPAAAS-----ADLTAAG----- 344
Qy 923 NHTLNKSDVITVLEQOQDMWFGVQOGKGMFPKSYVVLISGPRKSTSIDTGPESPAS 982

```

Db 345 -----DV----- 346
 QY 983 LKRVASPAKKAIPGEEFIAMTYESSHQGLTFQOGGVIVVTCKDDMMTGTVGDSGV 1042
 Db 347 -----EYIIAIPYIESAEGLSFSAGEMVVIKKEGEMTGTIGSTGM 391
 QY 1043 FPSNVYRLKD----- 1052
 Db 392 FPSNVYQADVGTASTAAEPVESLDQETTLNGNAAYTAPAVEAQEQVQPLPQVBPSEQ 451
 QY 1053 ---SEGS-----TACT-----GSLKKEPIA 1072
 Db 452 PISPGVGAEEAHEDLDTEVSQINQSTQSSBPAPETSRPMSRTSSMTPRAKRSEIA 511
 QY 1073 QVIASVATGPEQLTLAAGQILIRKKNPGWEGELQARGKRRQIGFPANVYKLLSPG 1132
 Db 512 QVIAPYEATSTEQSLTGTQGLIMTRKTDGSGWEGELQAKGRRQIGFPATYVYKVLQGG 571
 QY 1133 --TSKITTELPKTAQVQAVC-QVIGMYDYTAQNDDELAEKQGIINVLNKEDPDMKGE 1189
 Db 572 RNSGNTFVSGSRIMETEQIILDKVIALYPRKQNDDELSPDKDIISVLGRDEBEMRGE 631
 QY 1190 VSGQGLFPSPNV 1202
 Db 632 LINGSLGFPSPNV 644

RESULT 6

Q9UBC2 PRELIMINARY; PRT; 864 AA.

AC Q9UBC2
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAY-2002 (Tremblrel. 20, Last annotation update)
 DE Epidermal growth factor receptor substrate Eps15r.
 GN Eps15r.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 ON [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Nakashima S., Morinaka K., Ikeda M., Kishida S., Koyama S.,
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Ueki N.;
 RL "HRI NTT human fetal brain cDNA Project."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF110265; AAF21930.1; -
 DR EMBL; AB015346; BAA88118.1; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_repeat.
 DR Pfam; PF00036; efhand; 4.
 DR SMART; SM00054; EFh_3.
 DR SMART; SM00027; EH_3.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
 KW Receptor.
 SQ SEQUENCE 864 AA; 94254 MW; FA126069F6E00387 CRC64;

Query Match 8.6%; Score 536.5; DB 4; Length 864;

Best Local Similarity 23.5%; Pred. No. 1.7e-17; Matches 216; Conservative 139; Mismatches 349; Indels 215; Gaps 28;

QY 15 WAITVEERAKHDOPLSLKPIAGFITGDQARNFFQSGLPQVLAQIUALADNMNDGRMD 74
 Db 121 WAWRVEERAKHDOPLSLKPIAGFITGDQARNFFQSGLPQVLAQIUALADNMNDGRMD 180
 QY 75 OVERSIAMKLIKLUQGIQVLPSTLPPVKKQOPVAISSAPAFGIGIASMPPLTAVAPVPM 134

Db 181 REFVAMHLYVRALEKEBPVSAALPPL-----IPSKRKTVFP 220
 QY 135 GSIPVNGSPPLVSSVPPAAVPLNAGAPVYIOLPLAFAPHAATWPKSS--SFSRSGGS 192
 Db 221 GAVPLPASP-----PKQSLASTSTSHGVS 246
 QY 193 QUNTUKIQAQSFVDVASAPPAEMAVPQSSRLKYROLFNSHDKTMSGHITGQARTLMQS 252
 Db 247 SLNSTGSLSPKSLKQTOPTVMVWVVPVADKMRFDEIFUKTLIDLIDGVYSGQEVKEIFMHS 306
 QY 253 SLPOAGLASIWNLSIDDOGKTLAEFLAMHLDIVAMS--GQPLPVLPPIYIPSPFRV 311
 Db 307 GLTONLAHITMALDTRQTKLSKQFALAMYFIOQKSKGIDPPQVLSPMVPPSESGT 366
 QY 312 RSGSGMSVSSSSVDQRLPEPSSSEDOQPEKKLPVTFEDKKRENFERSVLEKRRQAL 371
 Db 367 PGPDSGSLGSGF-----TGVELELDISQ-- 391
 QY 372 LEOQRKEQRLAQLBRADEKEREQROEAQKQLELEKQLEKQLEKQLEKQLEKQLE 431
 Db 392 -----ELAQLOREKYSLEQDIREKEFAIRQTSVQ--ELQNDLDRFTSSLOELEAQ 441
 QY 432 RREAAKR--ELERQOLEMERNRROELINORNKEQEGTVV--LKARKTLEFLEALND 486
 Db 442 KQDQDRIDEMDQOK-----AKLRMLSDVQKQCODETOMISLKTQIQOSQESDLKQGED 496
 QY 487 KKHQLEGLKQDIRCRLARQOEIEBTKSRRLRLAETHLQOQLOESQOMGRLLPEKQI 546
 Db 497 DLNRAKSELNRLQOEETOLEQSIQARVQLETTIKSLKSTDEINQASKLSQLESQOE 556
 QY 547 LSPQKQVQVQ--NSLHROSILTLKRALFAKELARQOLEQLEDEVERETRSKLOEIDVFN 604
 Db 557 AHRSLQYQVYLDGAHAGSLTDLANLSECVSLAERKSGAND-----DPKFN 603
 QY 605 QULKELEIHSKQLOKQSLAARLKQKEQERKSLLEKQEDAKRQVQERDKQMLEHVQ 664
 Db 604 --KALLFSNNQELHPDP-----QTEDPFKS--DPKGAADPFKGDPFQNDP-----F 647
 QY 665 QEQQPRKPHEDRLKEDSVRKKEAEERAKEMQD-----KQSLFHNQEPK 715
 Db 648 AQQTSTDPFGGDPKESDPRGSATDPFKKQKNDPFTSDPTTKPSSL-----PSK 701
 QY 716 LATQAPWSTTEKGPLTISQESVKVYVYALYPPESRSDELITTOGDIWVDESQTEP 775
 Db 702 L---DPFESSD--PFSSSVSSKSGSDPFTLDPFGSGSFNSAE-----GFADPFQWQKP 750
 QY 776 -----GWLGEELKGTGWFPAHYAEKI PENEVFTPAKPVYTLTSAAPKIALRETPAP 828
 Db 751 PPGPFTSSLAG-----AGPSDDPFKSKQDTPALP-----PKKAP-----PRP 789
 QY 829 LPTVTSSEPTTNMNADEFSTWSSSNEKPE-----TDNMDTMAQOSLTVPSAGQLRQ 882
 Db 790 KPQSGKSTVSQLGSHDP-----EAPDFQGLADSGDPFGSKGFGDPFGS----- 837
 QY 883 RSAFTPATATGSSPSPVLG 901
 Db 838 KDPFVPSAAXPKSASAG 856

RESULT 7

Q8WQ61 PRELIMINARY; PRT; 1253 AA.

AC Q8WQ61
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Eps-15 protein.
 GN Eps-15.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

```

0Y 5 PTPFGSSJLD-VWALIVEERAKHDQFJSKLJAG IIGDQAKNE FFOSGJQVUVAOWA 6
Db 116 PVASGVANWGMSIGVIDRLKEYOEFESIHPSNMCLPGNKVKGVLMOSKJPMSTILGTIWD 175
0Y 64 LADMANNDGRMDQVEFSIATMKILKILKQCYQULPSTLPMPMKQPAISASPAFGIGIASM 123
Db 176 LADDDKGNLDMHEFFVAMHLVYQTLQKRITIPSVLPELRKP-----GGAGP 222
0Y 124 PPLTAVAVPVMGSIPIVVGMSPELVSS-----VPPAAVPELANGAPVI 166
Db 223 PPKRAMPPPPGGA-----AMPRAPSGGFBGDGFVANFPKDIAIPPAIPIPLVAVVPMPT 276
0Y 167 QPLPAFAHAPAAATWPKSSSFSSRSGBSQULNTKLOKASFVNASAP-PAEAWVAPQSSRLK 224
Db 277 RIPPV-----GAVSSQGLIQTDPILPIGAVPMANADWVTPADLKR 317
0Y 225 YRQLFNSHDKTMSGHITGPCQARTILMOSSLFQAOULASITWLSIDIDQGLTAEFFILAMH 284
Db 318 FEELIFROSDLDKDGIVSGLEVKYDIFIKSIGIPQSRSLADIWALCDTNOGSKLTVEBOFALAMW 377
0Y 285 LIDVAMSGQPLPVLPVPEYIIPESPRRNVSSGMSVSISSSVQDQRLPEPSESDQDQBEK 344
Db 378 FVERKQROVDPHPVLNANMVPSPSMKATVAGVDL-----QOQEV 415
0Y 345 LPVTFEDKKRENFERSGVELEKKRQALLQOQRKEOERLAOLERAEOERKEEROEOEAKR 404
Db 416 KP-TYSNPELEMISKEITELRER-----RVLETFIQKEADVIRKAGEVRSL 466
0Y 405 QLELE-----KOLEKORELOREREERKEIERREAA-----KRELEROLEMER 450
Db 463 QSELDLTVALIKOLENONGGAKRIDLQAOVSHNTAYLANVSLDISNTNOQVTKINDOC 522
0Y 451 NRRQELLNQRNKEQEGIVVLKARKRTLEFELIALNDKKHQLEGKLODIRCLATQROEIE 510
Db 523 HMOEVTIN-----EOEGE--LNAKRS-----ELQKLTDEASLOKEVDSNNRELSKLTINHLQ 572
0Y 511 STNSRREIRIAIEITHLOOQLOESQMLGRLLRPEKQILSDQKQVQOQNSIHHDSILTLTKRA 570
Db 573 AT-----QOQISSVSRMVTQLELTOQM-----TDALLICRAA 606
0Y 571 LEAKELARQOLREQULDEVERETRSLQELDIVFNNOULKEIRIHSKOOLOKORSLEAARLK 630
Db 606 ME-----NONMEL-----VEYQUKIEPDDDEA--- 626
0Y 631 QKQOERKSL--ELEKQKEDAORVQOERDKOMLEHVQOEOERPPKPHEDRLKREDSVRK 686
Db 629 -----RKTLTKQOVQPKOD-----PFEEN-----NS 646
0Y 689 KEAEERAAPEMDDKQSRLLPHQPERA-----KLATQAPMSTT 722
Db 650 GAANQATNGFGSDPFSG--QPVNKAIAISTGPDPSFNMSGFDSGDFAGDQAGASARFQOT 707
0Y 726 EKGPLTISAQESVVKVYVYRALYPPFSRSHDEITIQG-----DIVWVDSQTEBPGLG 779

```

D6	QDDEFGSA-----FAANKSNAITPEPKDDGSGSPFALHAPTC-----	748
OY	780 GELKGGTGMFPANAAYAEKIPENEVPTPAKVVDLTLSARPKLALAEPLAVTSSEF----	836
D6	749 GQVLSPN---AOKSGPPRPDESPPALP-----FKKSXVPPEPPAPPRAAQPTGG	795
OY	837 --STPNWMADE-----SSTWSSSEKEPETDNMDTMAAOPLATVP-----	875
D6	796 FGSGGGGGFADFDPDFDKNLHIIPASPSATAIABP-----LPPLTPAPIPVSSGS	846
OY	876 -----SAGOLRORSAPTP-----ATATGSSPSPLV-----GCGE	904
D6	847 LLDSPFLTFDDDGQHKKQAASPNTPIVTPTHTLLQTSLSTPAAPSALASLSTSGSG-	905
OY	905 KYEGIQIA-----OALTYPRAKKNDHLNFKNKDVIYTLLEQDDMMWF	944
D6	906 SVAGAGAILPSSVTTITTPASLNNOHLSNSNPLOQRADVKLETAKVASVFDA-----F	960
OY	945 GEVGOGCKMFWKSYVKLISGFVRKSTSIIDTGP-----TESPASLKRVASPAKPAPIGEE	999
D6	961 GEIGTRKAFTP---SLITGP---TDPKDDPFKDYREDFPSIK--DPFAE---EGEE	1006

RESULT 8
Q60902 PRELIMINARY; PRT; 907 AA.

```

AC Q60902;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Eps15r protein.
CN Eps15-RS OR Eps15R.
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB;
RX MEDLINE=96003812; PubMed=7568168;
RA Wong W.T., Schunacher C., Salcini A.E., Romano A., Castagnino P.,
RT Pellicci P.G., Di Fiore P.;
RT "A protein-binding domain, EH, identified in the receptor tyrosine
RT kinase substrate Eps15 and conserved in evolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9530-9534(1995).
DR EPMC; U29156; AAA87202.1; -.
DR MGD; MGI:104582; Eps15-rs.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_repeat.
DR InterPro; IPR003903; UTM.
DR Pfam; PF00036; ehfand; 4.
DR SMART; SMO0054; EFh; 3.
DR SMART; SMO0027; EH; 3.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
SQ SEQUENCE 907 AA; 59320 MW; 2FA79DFBA8C834BE CRC64;
```

Query Match 8.4%; Score 527, DB 11; Length 907;
Best Local Similarity 22.7%; Pred. No. 5e-17;
Matches 225; Conservative 134; Mismatches 362; Indels 270; Gaps 30;

OY	15 WAITTEERAKKHDOQLSLKPIAGFITGOANFFQSGLRPOVLAQIVMALDMNNDGRMD	74
D6	121 WAVEREERAKRDGFIESTLIPVNGLLSGDKVKEVLMNSKLPLDLGRVWDLSIDKDHLD	180
OY	75 QVEFSIAKKLIKTKLGQGLDPTSLPPWKQOPVAISSAPAFGIGIASMPPLTAVALPYPM	134
D6	161 RDEFVAHHILTVRALEKEKPPVISILPPPL-----IPSKRKKTIFA	220
OY	135 GISPVGVMSPLLVSVPAAVPLLANGARPVIQPLPAFAHPAATVPKXS--SFSSSGPGS	192
D6	221 GAIVPLPASP-----PKDSLSTPSHGGSV	246
OY	193 QUNTLCQAQSNDVASAPRAEMAWPQSSRLKYRDLFNFSHDKMTSGHLTGPARILLMQS	252

```

Db      SLNSTRGSPKHSV-KQPPVA-WVVPVADKKRPDEIFLKTDLDDGYSGGEVKEIFMHS 304
Qy      SLTPQALASIMNLSIDDDGKLTAEFFLTMHLIDVAMS-QQPLPVLPEVYIPSPRRV 311
Db      GLTQWLLMHLALADRTGTGKLSKEQFALAMYFIQKYSKIDPPQVSPVMPVPSRG 364
Qy      RSGSGSVISSSSVDQRLPEEPSSSEDOQPEKLPVTFEDKKRENFERGSELEKRRQAL 371
Db      PIPSSSTLAGEF-----TGVELEDDISQ-- 389
Qy      372 LEQAKREORLAQLEFRAGEKREKREKREKQLEKQLEKQLEKREKREKREKREK 431
Db      390 -----EIAQLOREKYSLEODIREKEEAIKQTSFVQ-----ELQNDL 427
Qy      432 RREAKRELEKROLEWERNRROELLNQRNKEQEGTVLKKARKTLFELEALNDKXHL 491
Db      428 RETSSQLELEKQKQDAQR-----LDMDQQA 457
Qy      492 EKKLQDLCRLATQROEISTNKSRELIATHLQQOQLOESQOMGLPEKQILSDQ 551
Db      458 RDLMSDVOKOCODETQTISLTKTQIOSQESDLKQEDDLNRAKSELNRLQGEF----TOL 513
Qy      552 KQVQONSILHRSLLTKALAKELARQLEQDDEVRETRSKLOEIDVFNNGKLKRE 611
Db      514 EQ-----SIOAGRAQLETTILRSUKTQDDINQARSKLSQLE 550
Qy      612 IHSKOLOKORSLE-----AARLKOKEERKSLLEKQEDAQORVQERKQML- 660
Db      551 SH-----LEAHSLEQYDVPDGVSTSLPDLATINEGILLAREGGFAMDDPFKNKALLF 606
Qy      661 EHVQOEOPRP--RKPIEDRLKREDSVRKKEAEERAKPEMOD-----KQSL 706
Db      607 SNNSELEHPDFOAEDPKSDPFKADPKADPFQSDPFSSQOTAYDPFGSDPFKESDP 666
Qy      707 FHPHGEPAKLATQAWSTTEKPLTISAQESKVVVYALVPFERSHDETTIQGIV- 765
Db      667 FHSSSDPFKQKTNDPFTSDPFTKNSLSKLDPFSSDPSSSSISKSGSPFGTLD 726
Qy      766 -----WDESGTGEF--GWLGEIKGKTGMFANVAEK-----IPENEVPTP 805
Db      727 PFGSSSFSSAEGFADPFSSOMSKRPSPGSPSSSL-GGTGFSDDPFKSKQDTPALPKK-PAP 784
Qy      806 AK-----PVDTLSAPAPKALARETP--APLPVTSSEPTTPNNMAD-SSSTMP- 851
Db      785 PRPKPSSQSTFVSQSGSDPF-----ESPDPQPLGADSGEPQONKKGFDPDPSGKDPF 839
Qy      852 -SSSNEKETDNDTWAQPSLTVPSAGLQORSAPFTPATGSSPSPVLQOGEKVEGLQ 910
Db      840 APSSSAK-----PKASS-----SGFADFTSGFQNEQQLAMAKRESEKAE 879
Qy      911 AQALYPMAPAKKDNHLNFKSDVITVLEQDM 941
Db      880 QERLARLRROEDL-----ELATALSKADM 905

```

RESULT 9

```

ID      Q23418      PRELIMINARY;      PRT;      751 AA.
AC      Q23418;
DT      01-NOV-1996 (TREMblrel. 01, Created)
DT      01-NOV-1996 (TREMblrel. 01, last sequence update)
DT      01-MAR-2002 (TREMblrel. 20, last annotation update)
DE      Hypothetical 81.8 kDa protein.
GN      ZK1248.3.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC      Rhabditidae; Peleoderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RX      MEDLINE=99069613; PubMed=9851916;

```

```

RA      None;
RT      "Genome sequence of the nematode C. elegans: a platform for
RT      investigating biology. The C. elegans Sequencing Consortium.";
RL      Science 282:2012-2018 (1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      latreille P.;
RL      "The sequence of C. elegans cosmid ZK1248.";
RN      [3]
RP      Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RC      Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RT      "Direct Submission.";
RT      Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
RL      EMBL; U29244; AAC71084.1; -.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR000261; EFh5_repeat.
DR      SMART; SM00036; EFhand; 3.
DR      SMART; SM00027; EH; 2.
DR      PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW      Hypothetical protein.
SQ      SEQUENCE 751 AA; 81849 MW; A21027D7AD8E8441 CRC64;

Query Match      8.1%; Score 506.5; DB 5; Length 751;
Best Local Similarity 24.7%; Pred. No. 3,7e-16;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;

Qy      15 MATTERAKHQOQFSLKPIAGFITGDOARFFPGSGLPQVLAQIWMALAMNNDGMD 74
Db      111 WPTTPDQAKYVSTFOSLNPVNGKLSGAHVRYLWNSGLDAHALAKIWEISDQDKGSLD 170
Qy      75 QVEFSIAMKLILKLYGYQLPSTLPP--VMKQPVAISSAPAFGIGIASMPRLTAVAV 132
Db      171 RIEMSVALLHYRSLQSDPVPALPPLNIHSEKAMVAHSSNF-----AAP--HPP 221
Qy      133 PM-----GSIPVGMSPPLVSVPPAAPPPLANGAPRVYIOLPARAHNA----- 176
Db      222 PMWGSAGSVTSLDDNNSSQ--YSATMPRAYP-----QPRATSAQAHVNGST 271
Qy      177 -ATWPKSSFSRSGSQSLNTKLQKAGSFVVASAPAEAMVPOSSRLYKQLFNSHDKT 235
Db      272 GASTPISASHS-----IHSF-----PAGEWPIPTGD--YADQFAQTDTN 308
Qy      236 MSGHLTGPAARTILMOSSLPQALASIMNLSIDDDGKLTAEFFLTMHLIDVAMSGPL 295
Db      309 KQGLVDMGMDRAPMTTGISAQILAHVALADIKKCGQLNLEQFALTMHLIDMAKRGSI 368
Qy      296 PVLPEVYIPSPFRVRSQSGSVISSSSVDQRLPEEPSSSEDOQPEKLPVTFEDKKRE 355
Db      369 PSELPHILIPSPFRPTESALHHPAQSVSTQPLQEAISMELKELF-----GE 417
Qy      356 NFERGSELEKRRQALLEQKREORLAQLEPAEGEREKREKREKQLEKQLEKQLEKQ 415
Db      418 NEF--MKQLAESIQMWLEKTAEBAVVLE-ADMTIKNSSIK--NIQVELATLESTV 470
Qy      416 RELEROREEERK-----EIERREAKRELEKROLEWERNRROELLNQRNKEQEGTVV 469
Db      471 KQERQKGEATRLADYDTQIEQLESA--CAQKETEEDTEKRYQIIDEKAKAED--- 524
Qy      470 LKARRLTLEFELEALNDK--HQLEGLQDLCRLATQROEISTNKSRELIATHLQ 527
Db      525 CKX-----NDEKEMELKKEIEMLDNQKRYTGEIVKTSQREQVAVLT--- 569
Qy      528 QOQLOESQOMGLRLEPKQILSDQLKOVQONSILHRSLLTKALAKELARQLEQD 586
Db      570 -----TIERKE-ARDQIQEMKLD 586
Qy      587 -EYERETRSKLOEIDVFNNGKLKREIHSKOLOKORSLEAARLKQKEERKSLLEKQ 645
Db      587 AATENTTKLTQVSDAVESSEEMQI-----LRSQQRLLSTIVIQSLSDDTVYGATG 641

```



```

QY 646 EDQRVRQERDKQWLEHVQOEOPRPRKHEDRLKREDSVRKKEAEERAKPEMOKQR 705
DB 642 TSSQ-----NHVQOPDP-----FASARANPAAD----- 665
QY 706 LFPHQEPAKLATQAPMSTTEKGLTISAQESVYVYRALYPFESRSHDEITIQPDIV 765
DB 666 -----PFAQVDQ-----FGSSGH----- 678
QY 766 MDESQTEGPEWGLGELKGTGWFPAN--YAEKIPENEVTPAKPVTDLTSAAPKTLAR 823
DB 679 -FDAFFPTDPAQGG-----FPSDGFQAQAPAK-----PAP-----RPAPKSAR 719
QY 824 ETPAPLPVTSS-EPTTPNMADFS 847
DB 720 ETPVNDPFAPSGOSTOPAGFADFA 744

RESULT 10
Q9BIF4 PRELIMINARY; PRT; 796 AA.
ID Q9BIF4;
AC Q9BIF4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE EHS-1.
GN EHS-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21376507; PubMed=11483962;
RA Salcini A.E., Hilliard M.A., Croce A., Arbucci S., Luzzi P.,
RA Tacchetti C., Danielli L., De Camilli P., Pelicci P.G., Di Fiore P.P.,
RA Bazzicalupo P.;
RT "The Eps15 C. elegans homologue EHS-1 is implicated in synaptic
RT vesicle recycling.";
RL Nat. Cell Biol. 3:755-760(2001).
DR EMBL; AY027560; AAK13051.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_Repeat.
DR Pfam; PF00036; ehnd; 3.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF_HAND; 5FAE9474EC258E98 CRC64;
SQ SEQUENCE 796 AA; 86644 MW; 5FAE9474EC258E98 CRC64;

Query Match 8.1%; Score 506.5; DB 5; Length 796;
Best Local Similarity 24.7%; Pred. No. 4e-16;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;

QY 15 WAITVERAKHDQFLSLKPIAGFITGDQARNFFQSGLPQPVLAQIWMALAMNNGRMD 74
DB 156 WPTPTDQAYDSIFOSLVNNGKLSGAHVRYVLMNSGDLAHALRIWELSDQDKGNLD 215
QY 75 QVEFSIAMKLKIKLQGYLPSTLP--VMKQOPVAISSAPAGIGIASMPPLTAVAV 132
DB 216 RIEMSVLHLVYNSLOSQDPVPAQLPRLILHPSKAMYAHSPN-----AAP--HPR 266
QY 133 PM-----GSIPVVGMPSPVSVPPAAVPLPANGAPPVIOPLFAFHPA----- 176
DB 267 PMNGSRAGSVTSIDVVMQS--YSATMPRAYP-----QPGAYSAQAHVNSRFS 316
QY 177 -ATWPPSSSSSRSGPSQNLTKLOKQSPVVASAPPAEWAHVQSSRLKYROLFNSHDT 235
DB 317 GASTPISASHS-----IHSF-----PAGEWPINTGD---YAOQFQDTN 353
QY 236 MSGHLTSPQARTILMOSLPQAOIASIWNLSIDODKLTAEFIILAMHLIDVAMGQPL 295
DB 354 KDLGVDMKRAKPMMTTGLSAQILAHVWALADIKKCGQLNLEQFALTMHLIDMAKGSST 413
QY 296 PVLAPPEYIPSPFRVRSVSGMSVSSSVDRLEPPESEDEQPEKKLPVTFEDKRE 355

```

```

DB 414 PSELPLHILPPSPRPPEPSALHHPQSVSTPQLPEATSMIEKALE-----GE 462
QY 356 NPERGSVELERKQALQEOQKEOERLAOLERABOEKREBEROEOAKROLELEKOLEQ 415
DB 463 NEE--MQLAESIOMLVERTKAEVQLQLE-ADMTIKNSIK-----NLQVELATLESTV 515
QY 416 RELEROREEERK-----EIERREAKRELEROROLEMERNNRROELLONKQEGSTV 469
DB 516 KOLERKEATRLRLADYDTQIEQLESA---CKAQKETREDETEKMQOIDDQAKAED--- 569
QY 470 LKARKLTLEFLEALNKK--HOLEGKLODRGLATOROIEBTSNRSRELTAEITHQ 527
DB 570 CRA-----NDEKEMELKEIEMLDNPFKTVRGIVYKTSORQKVAELT--- 614
QY 528 QOLOEQMLGRLIPEKQILSDQKOVQONSILHRDSLTLTKALAEAKELARQOLR-EQLD 586
DB 615 -----TLERKE-ARDQIQMKLD 631
QY 587 -EVERETRSKLOEIDVFNQKLEIREIHSKQOLOKQSLERARIKQEQERKSILEKQ 645
DB 632 AAIETTTKLTQVSDAVEKSEEMVQI-----LRSQORLSTVIDQSLSDDTVYGETAG 686
QY 646 EDQRVRQERDKQWLEHVQOEOPRPRKHEDRLKREDSVRKKEAEERAKPEMOKQR 705
DB 687 TSSQ-----NHVQOPDP-----FASARANPAAD----- 710
QY 706 LFPHQEPAKLATQAPMSTTEKGLTISAQESVYVYRALYPFESRSHDEITIQPDIV 765
DB 711 -----PFAQVDQ-----FGSSGH----- 723
QY 766 MDESQTEGPEWGLGELKGTGWFPAN--YAEKIPENEVTPAKPVTDLTSAAPKTLAR 823
DB 724 -FDAFFPTDPAQGG-----FPSDGFQAQAPAK-----PAP-----RPAPKSAR 764
QY 824 ETPAPLPVTSS-EPTTPNMADFS 847
DB 765 ETPVNDPFAPSGOSTOPAGFADFA 789

RESULT 11
Q91WH8 PRELIMINARY; PRT; 599 AA.
ID Q91WH8;
AC Q91WH8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to epidermal growth factor receptor pathway substrate 15,
DE related sequence.
GN Eps15-RS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=KIDNEY;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015259; AAH15259.1; -.
DR MGD; MGI:104582; Eps15-rs.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_repeat.
DR Pfam; PF00036; ehnd; 4.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 599 AA; 66206 MW; B62762A063F21FA9 CRC64;

Query Match 7.9%; Score 493.5; DB 11; Length 599;
Best Local Similarity 24.8%; Pred. No. 1.2e-15;
Matches 152; Conservative 96; Mismatches 189; Indels 177; Gaps 13;

QY 15 WAITVERAKHDQFLSLKPIAGFITGDQARNFFQSGLPQPVLAQIWMALAMNNGRMD 74

```

```

DB 121 WAVEEKAFCDFGFIESTLIPVNGILSGDKVPLVMSKSLDLVDLIGVWDSDIDKXGHL 180
QY 75 QVESIMXKILKLGQYOLPSTLPPVMOQOPAISSAPAFGIGIASMPPLAVAPVPE 134
DB 181 RDEFAVAMHLYVRALEKEPVPSTLPPV-----IPSRKKTIVA 220
QY 135 GSIFVGMSPPLVSSVPPAAVPLANGAPVLOPLPAFAHPAATWPKS--SFSRSGPGS 192
DB 221 GAVFVLASPP-----PDSLRSPSHGVS 246
QY 193 QLNTKLOAKOSFVNASAPPAEAVPOSSRLKROLFNSHDKTMSGHLTPQARTIMOS 252
DB 247 SLNSTGSLSPKHSV-KOPVVA-WVPVADKRRFDEITLKTDLDDGVSQGVKEIFMHS 304
QY 253 SLPPAOLASIMNLSIDODGKLTAEFFILAMHLIDVAMS-GQPLPVLPEVYPPSEFRV 311
DB 305 GLTONLAIHIALADTQGTGLSKGOFALAMVFIQKVSXGIDPQVLSDFWPPSEGR 364
QY 312 RSGGMSVSSSVQDRLPEPSSSEDOQPKKLPVTFEDKRENERGSGVELEKRRQAL 371
DB 365 PIPSSSTLASGEF-----TGVELDDISQ-- 389
QY 372 LEQARKOEEROLERAERKEREROEAKROLLEKOLEKORLEERQREERKEIE 431
DB 390 -----EIAQLQREKYSLEQDIREKEAIIKQKTEVQ-----ELQNDLD 427
QY 432 RFEAKKELEEROLEWERNRROELNQRNKEOEQTVLAKARKTLEFEELANDKKHOL 491
DB 428 RETSLOLEAKQKQDAQR-----IDEMOQKAKL 457
QY 492 EKGQDTRCRLATOROELESTNKREBLIAITHLQOQLQESQOMLRLLPEKQIISDL 551
DB 458 RDMLSVDRQKQDPTQTITSLKTQIQESDLSKQEDLNRAKSELRLQOE---TOL 513
QY 552 KOVQONSILHRDLSLTLKALAKELARQOLREQLDEVERETRSKQIDVFNQKLKLE 611
DB 514 EQ-----SIQAGRAQLLETLIRSLKCTQODINQASRLSQLOE 550
QY 612 IHSKOQLOKORSLE 625
DB 551 SH-----LEAHRSL 560

RESULT 12
QY 09W11 PRELIMINARY; PRT; 668 AA.
AC 09W11;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG16932 protein
GN ERS-15 OR CG16932.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Tracheata; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman U.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller J.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

```

RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howard T.O., Wei M.-H., Idegum C.,
RA Jatali M., Kallush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mikhaylov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtk R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003465; AAF47267.1;
DR Flybase; FBgn0035060; Eps-15.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002061; Eps15_repeat.
DR Pfam; PF00036; efhand; 4.
DR SMART; SM00054; Efb; 2.
DR SMART; SM00027; EH; 3.
DR PROSITE; PS00018; EF_HAND; 2.
DR Calcium-binding.
SQ SEQUENCE 668 AA; 73275 MW; 1D896D2431A27A2E CRC64;

Query Match 7.7%; Score 485; DB 5; Length 668;
Best Local Similarity 25.4%; Pred. No. 3,3e-15;
Matches 162; Conservative 104; Mismatches 224; Indels 148; Gaps 20;

QY 5 PTFPGSLD-VWATVERAKHDQFLSKPLIAGFITTDQARNFFPQGLPQVLAQIWA 63
DB 116 PVASGVANGDWSIGVIDRLKYEQLFESLHPSNGMLPNKVKYGLVMSKSLPMSILGTWD 175
QY 64 LADNNNDGMDQVESIMXKILKLGQYOLPSTLPPVMOQOPAISSAPAFGIGIASM 123
DB 176 LADQDQDGLMDHEFVVAHNLVYQTLQKRTIPSVLPPLKRP-----GAGP 222
QY 124 PLTVAVAPVPMGSIPIVGMSPPLVSS-----VPPAAVPLANGAPVVI 166
DB 223 PPKAMPPEPPAGA-----AMPRAPSGEGFGDGFVANFPKDIAPPAIPLPAVAPVMT 276
QY 167 QPLPAFAHAAVWPSSSFSRSGPSSQANTLKQAKOSFVNASAP--PAEAAVPOSSSLK 224
DB 277 RIIPV-----GAVSSQPLIQTDPLPIGAPMANAVANVPPADIKR 317
QY 225 YROLFNSHDKTMSGHLTGPOARTIIMOSLPPAOLASIMNLSIDODGKLTAEFFILAMH 284
DB 318 FEFIFQSDLDGDLGVSGLVXKDIIFKSGIIPORSLADIALCTNQSGLTVEQFALAMW 377
QY 285 LIDVAMSGQPLPVLPEVYPPSEFRVRSVSGMSVSSSVQDRLPEPSSSEDOQPKK 344
DB 378 FVERKQRGVDPVHVLNANVPPSMKATVAGVDL-----QPEV 415
QY 345 LPVTFEDKRENERGSGVELEKRRQALLQOEERLAOLERAERKER-EROEOEAK 403
DB 416 KP-TYSPLEMIISKEIELARER-----RVLEFIAKEDADVAIKNGESLD 462
QY 404 RQLELEKOLEKRELEERQREERKEIERREA-----KRELEROLEWERNRROE 455
DB 463 TLATATIKQLENRGAEKQKRLDLQAGVSHNTAVLANVSLDSRTNEQVYTKIRDCQHQEV 522

```

```

QY 456 LNMORKEOGTVLVKARKTLEFELALNDKKHOLEGKODIRCLATQROESTNKS 515
DB 523 TIN-----EOGE--LNKRS-----ELOKLEDEASLOKEYSDNNRELSTKTNHLOAT--- 569
QY 516 RELRIAEITHLOOQLOESQOMGLRIPERKQISDOLKOVQNSLHSDILTLKRALE--- 572
DB 570 -OLOISSVRSMVTOLETOROM-----TDALLICRAAMENON 605
QY 573 AKELARQOLEOLD-EVERETRSK---LOEIDVF--NN 604
DB 606 AELVSEYOLKIEPDPDEARKTTLKEVOLPKDDEPEENN 643

RESULT 13
Q9HGL2 PRELIMINARY; PRT; 1116 AA.
ID 09HGL2;
AC 09HGL2;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 120.5 kDa protein.
GN SPBC800.10C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McCombie W.R., Lyne M.;
RT "Sequence analysis of a region of the fission yeast genome.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McCombie W., Lodhi M., Kaplan N., Johnson A.;
RT "Sequence of a region of the fission yeast genome.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391034; CAC01525.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00261; EFS1_repeat.
DR Pfam; PF00036; ehand; 4.
DR SMART; SM00054; EFh; 3.
DR SMART; SM00027; EH; 3.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KM Hypothetical protein.
SQ SEQUENCE 1116 AA; 120540 MW; 79F1941C4CEBAF9 CRC64;

Query Match 6.9%; Score 434.5; DB 3; Length 1116;
Best Local Similarity 21.0%; Pred. No. 1.4e-12;
Matches 215; Conservative 152; Mismatches 320; Indels 339; Gaps 39;

```

```

QY 301 -----PEYIP-----PSFRVRSGSGMSVSSSVYD----- 326
DB 355 VPTSPISVVPANISSPNPNPTLAPNPTGSPSVTSGTELDLSDATPPSPPLAPOHTSSNA 414
QY 327 -----QRLPEPSSDEDOQ-----PEKLPVTFEDKKREN 356
DB 415 TKHSAPVTKSAPPPVVSPLQNHPTGFPPTSPPAKPNSPTSTFPQSSFGOTIAKNTMDK 474
QY 357 FERGSELEKEROALLBQQRKEOERLAQLEBAOERKEREREOEAROLEKOLEKOR 416
DB 475 PSAVTSPSQLAAPIPVASAE---QKLAAVPRKLESOLGYKKSNDLQK---SSR 527
QY 417 ELERQEEERREIERREPAARELERQOLEMERNRQELLNORNKQEGTVLKARKKT 476
DB 528 DVAAVLSDVKAKVSEIRAYDEBLAKAKQI-----S 558
QY 477 LEFELALNDKKHOLEGKODIRCLATQROEISTNKSRELRIAEITHLOOQLOESQOM 536
DB 559 LDIE-----TNKAQTOVNRREVSILBATNALOKOKORKEV 595
QY 537 LGRLTPEKQILSDOLKOVQNSLHSDILTLKRALLEAKELARQOLEOLD-EVERETRSKL 596
DB 556 LEQVAAE-----SEAAKNVSS--NASTIOLKSEVADKEOTLAQHLQLDENTQ---RL 645
QY 597 QEIDVFNQLEKLEBIHKSQOLQORSLAARLKQKEQERKSLERKQK---EDAQRV 652
DB 646 VSLDESKAVSQ-RKLDLEVYKINSKQGLATATBEYHSHSQLEAKQELSKLEDGLKSV 704
QY 653 QERKQMLEHVQOEOPPRKPHEDRLKEDSVRKKEAEERAKPEMODKOSRLFHH-- 710
DB 705 N-----LTBEAPKP---EVDSTPRPSP--TSNGITTDKPLPDITSSVPTQHS 749
QY 711 -----QEPAKLA--TQAPWSTTEKCP--LTISA-----QESVKVYVYALY 747
DB 750 FDMAMNTLRSPSLNSNNSAASAVSRNPFNLKISGASSPVSFWSESPASAVFPSPIS 809
QY 748 PFESRSHDEITIQPQ-----DIVMDV---ESQTEPGWLGELGKT 786
DB 810 KTTLSLVNNSVNSVNSLOSEPVLQNMEEPHQSDSVVDVTSASQORSP--VLSDLKLT 867
QY 787 GWFPAIVAEKIPENEVPTPAKPVTDLTSAAP-KLALRE-----TPAPL 829
DB 868 G-SANTAEPV-EN---TSAPPIEN--TSAPPIEIAKQATBIPISAPFATETISTAPV 921
QY 830 --PVTSSP-----PSTPPNMADFSSTW-----PSSSEKPEPTDMMTMAAP 870
DB 922 KPPVPPSRDRSADQGVVQATPHIQDEFPPIQNEIDDESSDEEPMSN---LSP 976
QY 871 SLTVPS 876
DB 977 QISQIS 982

RESULT 14
Q18137 PRELIMINARY; PRT; 944 AA.
ID 018137;
AC 018137;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Hypothetical 113.6 kDa protein.
GN C25A11.4B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL_N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
"Genome sequence of the nematode C. elegans: a platform for

```


QY 387 RAQERKERERQO-----EAKQULELEKOLEKQERERERKEIERREA 436
 Db 333 RLIRQOREKREEREDRLBSIRLAEBAELARRALEKER-IDREKAEERKTMERLERE 391
 QY 437 KRELEROROLEWERNRRO---ELNQRNKEOGTV---VLKARKTLEFEEALNDKCH 489
 Db 392 RARLERER-LEERRROKEREKETERIERERREHRIEIERIKERIEREREREKKA 450
 QY 490 QLEGLQDIRCLATQROEIESTNKS-RELRIAEITHLOOQLOESQOMLGRLIPEKQILS 548
 Db 451 EED-----RLRRLERLERIERERRELEAREQOLELQRRREADR-----ERQRL 496
 QY 549 DQUKQY-QQNSLHRDSULTIKRLEKELARQOLEDEVERETSKLOEIDVFNNU- 606
 Db 497 DEAREMRRREREERREAEVADVHQAEEERERLKRQERBEAERLERIRLEQOKIDMERID 556
 QY 607 KELREIHSKOOLQOKORSU-EAARLKQEOERKSU---ELEKQEDARVRVOERDKOWLEH 662
 Db 557 AERREERERKEERERREFELI EAARKKEARDRDLDEMERERVAREERERKEERERERI 616
 QY 663 VQOEOPRPRKPHIE-----DRLKREDSYRKKEAEERAKPEMO 700
 Db 617 AAEKERKROEEREEELARLNELORAAARQAQONNAELDRQORQDELDRK-AQELSEREMR 675
 QY 701 DKQSRLEPHQEPKATQAPWSTTEKGPLTISAQESVKVYTRALYPFESRSHDEITQ 760
 Db 676 EKERRDRERANBEAQLADLERERHNUIRENERREAVERRANNRLE--DRSRDCL--- 730
 QY 761 PGDIVVDESQTEGPEWLGELGKGTGWFPANYAEKIPENEVPTPAKPYTDLTSAPAKL 820
 Db 731 --DHIVERSEKEQ-----FELEKRLLAKEKANNRKK--NHLISSE---- 768
 QY 821 ALRETPAPLPLVTSSEPSTP-----NNWADFSSTWPSSS 854
 Db 769 TLAKLTQPMYTTREBPVTTKVERQYIERIDRNVWVEDVPYAPSQS 814

Search completed: March 14, 2003, 12:09:44
 Job time : 51 secs

